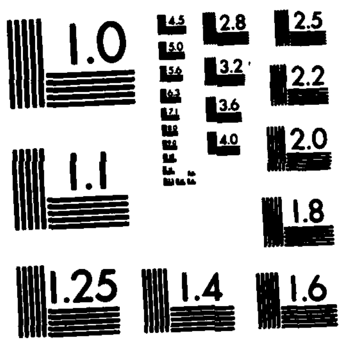
AD-A124 194 PROJECT ANALYSIS GUIDE(U) OPERATIONAL TEST AND EVALUATION FORCE NORFOLK VA 05 FEB 81 COMOPTEVFOR-INST-3960.8 1/2

UNCLASSIFIED F/G 5/2 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AD A124194

PROJECT ANALYSIS GUIDE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO. AD-A124194	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Project Analysis Guide		5. TYPE OF REPORT & PERIOD COVERED Handbook
		6. PERFORMING ORG. REPORT NUMBER COMOPTEVFORINST-3960.8
7. AUTHOR(s) Commander Operational Test and Evaluation Force		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of the Navy Commander Operational Test and Evaluation Force Norfolk, VA 23511		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS		12. REPORT DATE 5 February 1981
		13. NUMBER OF PAGES 75
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

This document has been approved for public release and sale; its distribution is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

Unlimited Distribution

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Test and Evaluation, Analysis, Experimental Design, Operational Test and Evaluation, Project Operations, Test Planning, Operational Suitability

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This document is designed primarily for the Operational Test Director (OTD) and provides guidance in test planning and experimental design for operational test and evaluation.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

DEPARTMENT OF THE NAVY
COMMANDER OPERATIONAL TEST AND EVALUATION FORCE
NORFOLK, VIRGINIA 23511

COMOPTEVFORINST 3960.8
02B:mas
05 February 1981

COMOPTEVFOR INSTRUCTION 3960.8

Subj: Project Analysis Guide

1. Purpose. This document provides guidance for various facets of analysis in OT&E (operational test and evaluation). It is designed primarily for OTDs/OTCs (Operational Test Directors/Coordinators) of COMOPTEVFOR Staff. Subordinate commands may supplement it as necessary according to their needs.

2. Future Changes

a. OT&E is a dynamic evolving process; suggested changes to this Guide are encouraged. Address them to Code 02.

b. Project Analysts will be asked to comment on the Guide's contents annually, in an effort to ensure continuing Guide utility.

H. A. French
H. A. FRENCH
Deputy Chief of Staff

Distribution:
(COMOPTEVFORINST 5216.2B)
List II, 1, 5

Copy to:
DEPCOMOPTEVFORPAC (50)
CO, VX-1/VX-4/VX-5 (50)
COMOPTEVFORDET Sunnyvale CA (10)
COMOPTEVFORDET Pax River MD (4)
VX-5 DET Whidbey Island WA (6)

DTIC COPY INSPECTED 2

Accession For	
NTIS GRA&I	X
DTIC TAB	☐
Unannounced	☐
Justification On form 50	
By Copy on file	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A	

Contents

Contents (Cont'd)

References

(a) Analyst's Notebook, COMOPTEVFOR INSTRUCTION 3960.7. This is a handbook of data analysis procedures to supplement standard analysis texts. The coverage is highly selected, non-mathematical, and straightforward. While the aim is to refresh the newly arrived analyst, the OTD may find certain portions of interest.

(b) M. G. Natrella, Experimental Statistics, National Bureau of Standards Handbook 91, U.S. Government Printing Office, Washington, DC, 1966. If the OTD is stuck without analysis help, this is recommended. The Table of Contents is descriptive as to the problems covered - comparing averages of two materials, etc. The analysis techniques are given in step fashion - in general terms and correspondingly with an illustration. This is an excellent cookbook.

(c) S. Siegel, Nonparametric Statistics For the Behavioral Sciences, McGraw-Hill, New York, 1956. This is an excellent cookbook for analysis of count type data (hits/misses, yes/no type). Each technique is illustrated with a worked-out example. The coverage of count type data analysis techniques is extensive.

Section 1

Introduction To Analysis

101. Analysis and Synthesis. In the broad sense, analysis is dividing the whole into smaller and smaller pieces until they become manageable for direct treatment. For example, in project planning the general objectives may lead to a subobjective to determine probability of kill, which in turn may include probability of detection as a sub-subobjective. An important phase of analysis is synthesis in the opposite direction -- combining pieces into larger and larger pieces. For example, the probability of detection that has been determined may be combined with results on classification, acquisition, etc., to derive a probability of kill that may in turn be combined with other measures to address a general project objective.

102. Planning and Data Analysis. Each analysis step in planning with finer and finer subdivisions has a counterpart in synthesis after project operations that leads to reporting of larger and larger combinations. (See Figure 1-1.) Thus, analysis in planning is directly related to data analysis in reporting. The words may not be the same, but the only real difference is the possible influence of surprises during project operations. This backs up the principle: The sooner the OTD visualizes the complete form and structure of the final report, the better the planning. The analysis and corresponding synthesis applies to project work at all stages. The measurement spectrum is an example. After testing, analysis is characterized, broadly speaking, by a continuous summary process. Thousands of measurements taken during testing

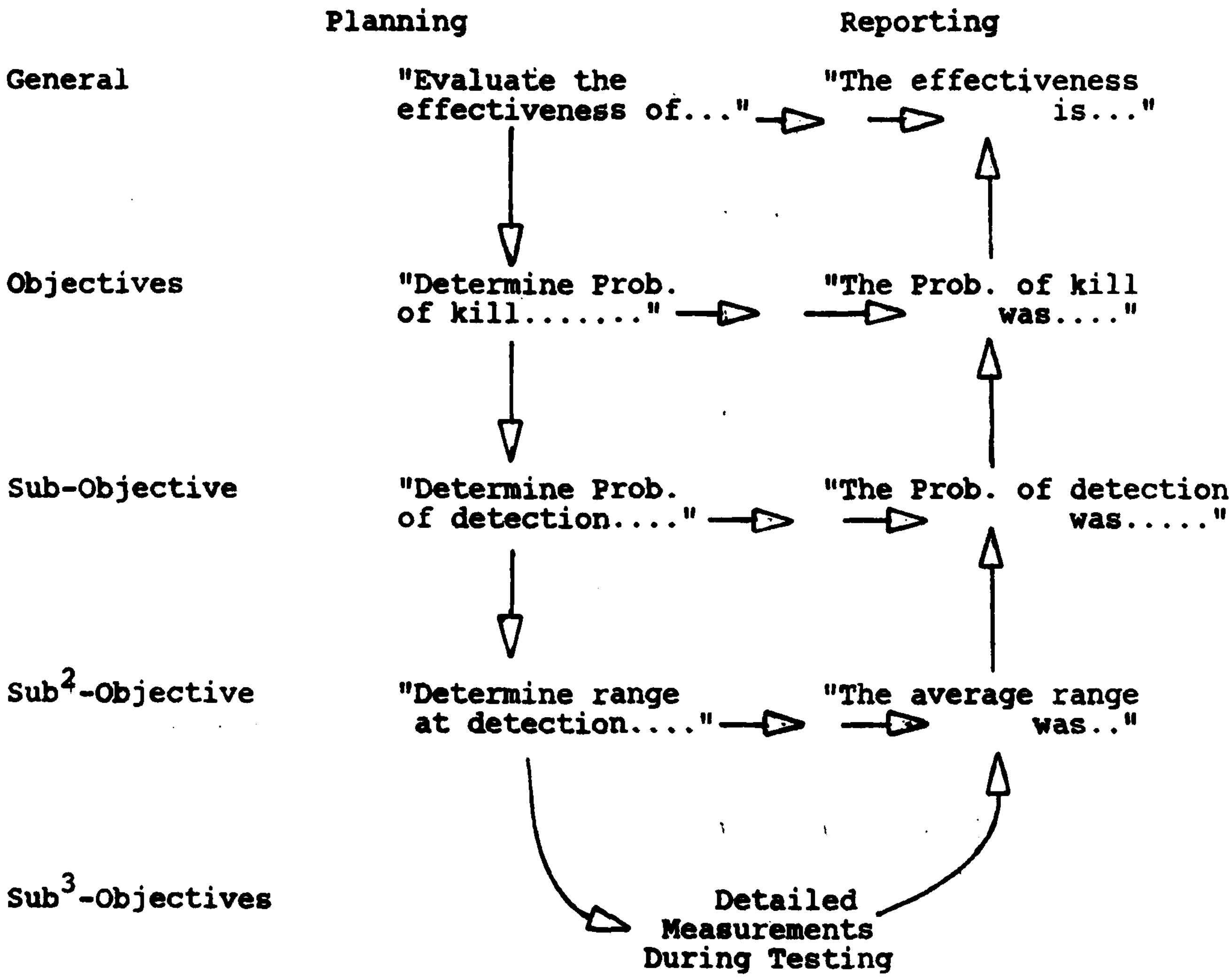


Figure 1-1

<u>Relationship Between Planning and Reporting</u>

are processed into data. Data are analyzed into performance measures or MOEs (measures of effectiveness). These are combined into a CEM (combat effectiveness measure), one of the end-products of analysis, an input into the evaluation process. The same spectrum is pertinent to project planning by going in the opposite direction. An important initial point is the CEM; then the performance measures (MOEs) are specified as components of the CEM. The next step is to determine the type of data, which in turn leads to the instrumentation needed to obtain such data from the planned tests. See Figure 1-2.

a. Measurements. In project operations, seldom can a performance quantity (e.g., detection range) be measured directly. To determine the detection range of a submarine sonar, for example, the positions of the sonar ship and the target are accurately tracked and recorded. This mass of measurement is correlated, and the range between the ship and target is found at the specific event, detection. This is done for each trial or run. The procedure of processing measurements into data -- data processing -- is usually the responsibility of a technical facility such as a test range. Our responsibility is to determine what we need, how much, how good, in what form, and when. Data processing is completed by preparation of a run-by-run summary, giving the test conditions, pertinent performance data, validity codes, and remarks.

b. Data

(1) Data are the result of processed measurements that include definition of a specific event. Thus, the distance between

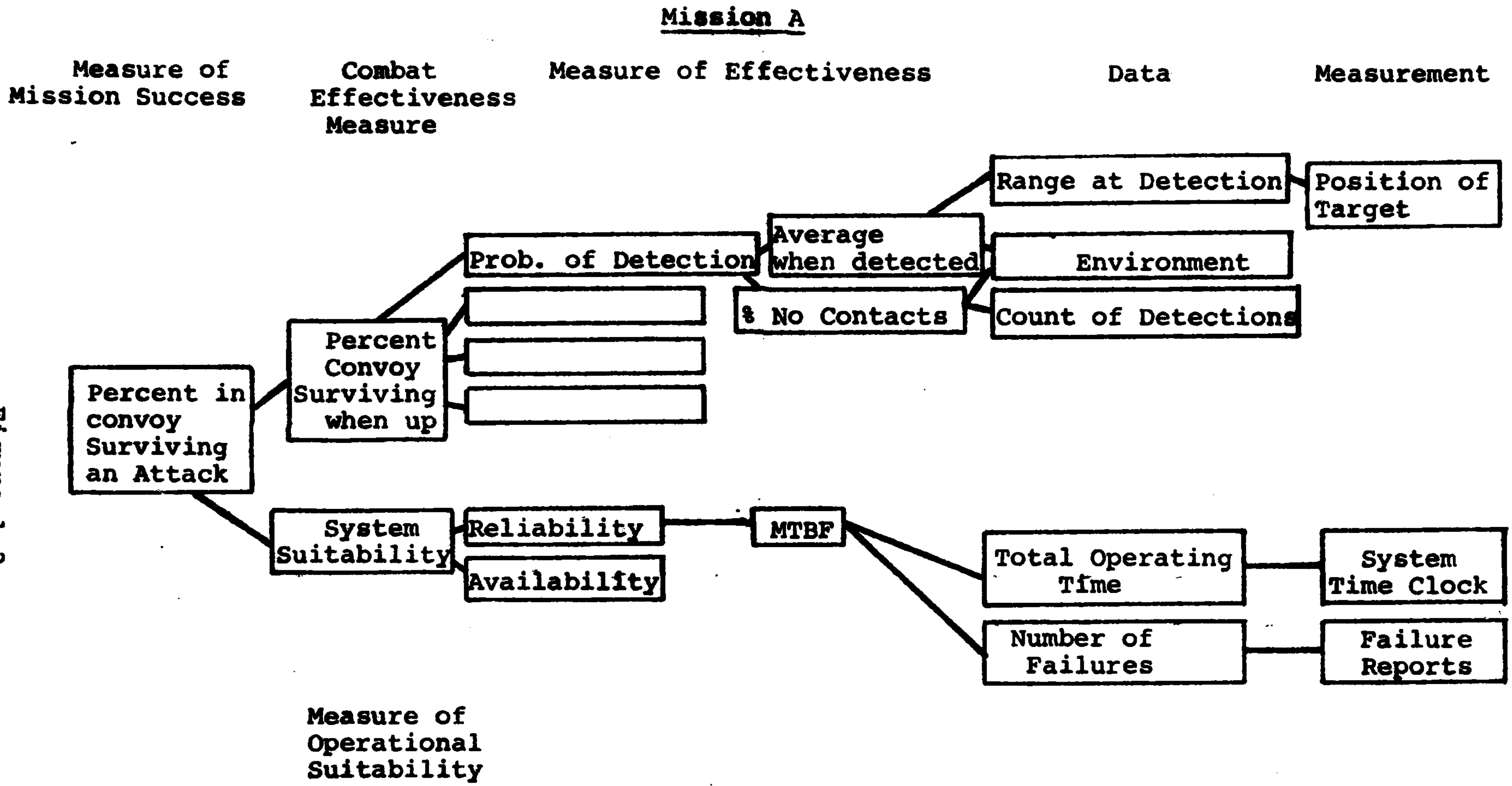


Figure 1-2

Illustration of Dendritic Structure

the sensor and the target at a specific event is taken as range of detection.

(2) Data are treated by data analysis; we analyze the number of hits, times to acquire, ranges at detection, etc. Each function such as detection or acquisition is analyzed separately. Basically, this analysis is to determine the effects of the different test conditions. This differentation sets limits to the summarization process -- should the data be summarized into one average, or should separate averages be reported for different conditions? The analysis and subsequent summaries provide information that is reported as Results. "The average range of detection for slow targets was 90 nmi; for fast targets it was 70 nmi."

(3) Initially, each type of data is analyzed separately. Then some types are combined. For example, a cumulative detection summary may include analysis of detection range and also no-contacts. Larger summaries or MOEs are made by combining summary measures into other MOEs. These summary measures differ in degree. Basically, they are characterized by combining more and more results into MOEs.

(4) Data can be continuous (e.g., the miss distance was 6.2 ft.), or can be categorized (e.g., hit, no-contact, failure). Continuous type data are usually expensive to obtain, involving a high order of instrumentation and data processing; to categorize a run outcome as hit or miss, success or failure, is cheap, relatively. However, more information is available in continuous type data. Knowing that detection occurred at 34,000 yds tells us more than knowing that detection has occurred. The added

information per run may result in the need for less runs. Thus, there is a tradeoff. We usually fund the instrumentation, etc., needed to get continuous type data; the payoff comes with reduced fleet service requirements.

(5) More and more in OT&E, qualitative data are becoming more critical. Qualitative data are frequently obtained through highly structured questionnaires filled in by well-qualified observers, operators, etc. These data may be supplemented by interviews and debriefings. With qualitative type data we try to get a consensus, so that the basis for results is as broad as possible.

(6) The accuracy needed in data has an important influence on the cost of instrumentation and data processing. Usually the accuracy needed in OT&E is gross compared to technical testing. However, in a project with a small sample size there may be increased need for instrumentation and data processing. For example, not all phenomena are probabilistic; not all events are statistical. A technical examination of an event such as no-detection or a hardware failure may indicate a technical cause or design deficiency. The review may indicate a deterministic situation, and firm results may be formed even with small sample sizes. So, as usual, there is a tradeoff as to need, cost, etc.

c. Measures of Effectiveness

(1) The classical definition of MOE is that it is a numerical measure of how well a task is done or an objective is met. For a sensor, the MOE may be the probability of detection. For a missile system, the MOEs may include the number of antiship

missiles that can be engaged successfully. For a surface ship, an MOE may be the likelihood of an escorted ship not coming under torpedo attack during an ocean crossing.

(2) The classical definition and use of MOE covers too broad a spectrum and may cause confusion. Hereafter, the term MOE will not be used per se, but will be modified. In a more narrow sense, we will use performance MOE or functional MOE. Later, for broad use, we will define CEM (combat effectiveness measure) and MOS (measure of operational suitability). Functional MOEs are summary measures based on data analysis (e.g., probability of detection) to describe a function (e.g., detection). Thus, the percentage of valid opportunities detected is presented as a measure of detection capability. It is standard practice to obtain these functional MOEs, particularly in comparisons of equipment. In some projects these functional MOEs may be sufficient as the end-product of analysis. In many projects, particularly involving systems, functional MOEs are not sufficient. They must be further combined as will be described later.

(3) A functional MOE depends on "other things being equal." If so, then a mine clearing equipment "A" with a sweep width (MOE) of 50 yd is better than equipment "B" whose sweep width (MOE) is 30 yd. Even so, this functional MOE comparison does not tell us what this improvement in sweeps width means operationally. Suppose equipment "A" requires a longer turnaround time -- then we would need a more complete measure of performance that included time. We would perhaps determine the time each equipment

took to clear a typical mine field. As you can see, we are now combining different functional measures: sweep width, time to sweep, clearance rate. To combine more and more functional MOEs, we have to "model" in an operational sense. We have to decide what is a typical mine field or fields; what is an acceptable standard level; what are the tactics to be used in mine sweeping; etc. The process or modeling becomes more and more complicated; the assumptions become more important; the calculations become more involved. However, we are obtaining more complete measures; these measures are coming closer to a measure of operational effectiveness -- a measure of the system's capability to do its job.

d. Combat Effectiveness Measure

(1) In OT&E, the classical interpretation of MOE is too narrow to be useful. In OT&E, the critical numerical value is how well a Naval unit will perform its operational mission pertinent to the system under evaluation in directly supporting the combat weapon system. This support concept is critical in defining operational MOEs. To stress this we use CEM, rather than the general or technical or non-operational MOEs.

(2) For example, an evaluation of a submarine sonar may indicate excellent detection performance. However, when it is interfaced with the submarine's fire control system, oscillations may be so prevalent that the firing solution is seriously degraded. The oscillations are not caused the sonar per se (which is being evaluated) but by the interface (which is not being evaluated). For scenarios involving torpedo firing, the sonar has poor operational effectiveness even though it has excellent technical

effectiveness. Its MOE is high but its CEM is low. The point is that the submarine commander would prefer the old sonar rather than the technically improved version as things stand. Thus, in evaluating a submarine sonar, the selected CEM (e.g., for a barrier-type mission, the likelihood of a penetrator being torpedo-engaged before the platform is engaged) may be low even though the fire control system is the culprit.

(3) Another example: a decoy device was very effective in decoying a torpedo when the decoy was properly launched from a submarine. However, the launcher on a particular class of submarine was found to be inadequate for launching this and all other devices. The decoy was not operationally effective for this submarine class until the launcher problem was corrected.

(4) CEMs are seldom measured or determined directly. Usually they are calculated, typically by combining functional MOEs. Thus, a mission/scenario may be decomposed into various functions. Functional MOEs relate to how well these functions are performed. For example, the CEM for an air-to-air missile may be an exchange ratio -- the probability of the target being destroyed before the missile aircraft is destroyed. The CEM could be determined by combining (e.g., by conditional probabilities) success measures of detection, positioning, launch, guidance, fuzing, and kill.

(5) There are many types of CEM -- the most complete measures of operational effectiveness pertain to exchange ratios. These ratios take platform survivability and attack effectiveness into account. Usually this measure is used for whole-ship or

aircraft evaluations. In evaluations involved with weapon systems, platform vulnerability may not specifically be taken into account. The most common measures used then are restricted to some form of kill probability. For a torpedo weapon system the "weapon system effectiveness" is the probability of kill of a single submarine or surface ship (one-on-one situation). For AAW the "saturation rate" pertains to the number of targets killed in a mass attack. In ASMD, a variation of probability of kill is used, pertaining to surviving.

(6) Here are some actual CEM results (camouflaged slightly for security reasons):

(a) In a self-protection situation under a wave attack at low altitude a TERRIER, double-ender, dual-launched ship at Condition I, has a saturation rate of eight targets. At high altitude it is 18 targets.

(b) The estimated probability of seaworthy impairment was about 0.25 for deep targets. For shallow targets the estimated probability of seaworthy impairment was lower and varied strongly by situation.

In the latter example, the "probability of kill" values were derived from "conditional values" using components of detection (P_d), location (P_l), aimpoint (P_a), missile delivery (P_m), and lethal radius (P_r). The latter component was obtained from a published reference.

$$CEM = P_d \cdot P_l \cdot P_a \cdot P_m \cdot P_r$$

Note that CEM pertains to performance assuming that the system is up in a materiel suitability sense.

e. <u>Measure of Operational Suitability</u>

(1) The system must be available for use when needed and continue to perform in the materiel reliability sense throughout its mission. MOS is usually related directly to operational availability. For a continuous-use type system the MOS may be simply the percentage of time that the operational commander can use the system compared to the total time needed during the mission. For expendable items such as missiles, the materiel reliability success ratio includes the likelihood of having a "good missile at launch" as well as the materiel success during attack. If repairs cannot be made during the mission, such as during an aircraft sortie, then the MOS treatment may be akin to that for expendable items.

(2) The analysis structure for performance applies also for reliability. Measurements include a breakdown of total test time into various categories; data include times to failure, times to repair; MOS includes mean time to failure, mean time to repair. These are finally combined into an overall MOS. As with MOEs, MOSs may have to be determined separately for various scenarios/ missions, etc.

(3) Usually a system includes a series of sequential functions or modes; the MOS may take the form of a product of conditional reliabilities.

$$MOS = R_1 \cdot R_2 \cdot R_3 \cdots$$

On the other hand, the system may include redundancy, turnaround, etc, and reliability modeling may be quite extensive.

f. <u>Measure of Mission Success</u>. CEM combined with availability and reliability (MOS) is called MOMS (measure of mission success). Then this combined measure is a more complete and realistic figure of what could be expected in an operational sense.

$$MOMS = CEM \cdot MOS$$

If the product of conditional values applies we have

$$MOMS = P_1 R_1 \cdot P_2 R_2 \cdot P_3 R_3 \cdots$$

Usually, however, a high order of modeling effort is involved. Even so, important elements will not be taken into account and have to be mentioned in "Limitations to Scope." For example, materiel reliability of the threat, vulnerability of the platform, etc.

g. <u>Elements of Essential Analysis</u>

(1) The dendritic or branching structure of decomposing an objective into subobjectives is a standard military planning technique. For example, the Army and Mitre Corp have implemented this into an EEA (element of essential analysis). An objective is structured into smaller and smaller pieces until manageable and specific. Then a subobjective is formed, called EEA. This is followed by CQs (checklist questions) that can be answered by data, yes/no, or a word.

(2) For example, Mitre used the EEA approach in AEGIS planning. An operational question or "critical issue" in effectiveness was divided into three branches. One of these branches was divided into seven limbs, one of which was: "What was the contribution of AEGIS to overall task force early warning?" This was divided into four EEAs -- e.g., "What percentage of first

detections was provided by AEGIS?" The corresponding CQ was "How many initial detections were made by each surveillance system?"

h. Methodology. The EEA approach and the MOMS approach are basically the same. However, use of each technique in its place can simplify planning. The methodology outlined in paragraph 215 combines both approaches.

(1) The CEM and MOS approach is used for quantitative elements.

(2) The "critical issue" approach and the EEA technique are used for qualitative elements.

i. Implementation. Although the dendritic analysis structure is simple and straightforward, its application to a specific project is usually complex, time-consuming, and more of an art than a science. For example, defining the functional MOE or CEM is only a part of the picture, and is often relatively insignificant in effort compared with implementation of the definitions. The complexity stems from the fact that the effort is not to summarize what happened during project operations per se. The aim is to predict the future outcome in a cold/hot war situation, when and where it will occur, against likely threats, etc. Obviously we can only try and take as many elements into account as feasible.

(1) In calculating CEM and MOS, the missions of the system and likely scenarios are paramount. For example, in ASW, will the system be used in a barrier or patrol, or both? Will we have prior intelligence? What type of targets? Will we have an alerted target? Will we be operating alone or in company?

These and many other questions have to be answered.

(2) The process of forming and using the scenarios, bringing the environment into the picture, determined target evasion, etc., involves a high order of skills and experience. Extensive modeling must be done. Usually a computer is involved. See paragraphs 203 and 604. In some fields involved with weapons systems, modeling has already been accomplished and models are available that are readily adaptable to our needs. In areas such as command and control and communications, we have to be content with a limited scope: functional MOEs.

(3) The emphasis in this analysis section is on quantitative measures. This does not imply that qualitative aspects are not important. Except for the obvious, there is not much that can be said about the qualitative aspects. See paragraph 214.c.

j. Terminology. The above terminology (MOE, CEM, MOS) is not standardized in the evaluation community. Others usually use only functional MOEs to cover the entire spectrum of measures. For example, a weapon system evaluation may have 30 different functional MOEs ranging from average detection range to probability of kill. These functional MOEs may or may not include reliability. Obviously it is important to know what is included in and what is excluded from each measure.

103. Scope of the Evaluation

a. The Navy's need for the system under evaluation includes a variety of aiming conditions, a variety of targets, a variety of installations, etc. Let's explore this further. If the system under evaluation is introduced into the fleet, it will be installed,

not only on our test ship, but on the many ships of its class. It will be used, not by our test crew on a four runs a day basis, but by a variety of crews under different levels of stress; not in a test range, but in the North Atlantic and the Med. The targets will not be a Guppy class, but a variety of classes with a variety of characteristics doing a variety of missions, maneuvers, and evasive tactics. We can build up an impressive list of varieties.

b. The important point is that our evaluation pertains to this variety. CNO is not interested per se in what happened last June in the warm waters of Key West with a special installation and an alerted crew. Our only interest in the happening last June is obtaining information on what will happen under the scope and variety of possible situations. Therefore, in our evaluation, it's better to cover as broad a scope as possible, rather than look in detail at a small portion. Our job is to sketch out the forest, not to describe a few trees in detail. How wide a scope should be strived for? Analytically speaking, the answer depends on the analysis. If we can't analyze the data, our results may be meaningless or misleading. There is a tradeoff that we will explore later.

104. Realism in the Evaluation. In addition to having a broad scope, our testing must strive for realism. Realism pertains to the expected use of the system under test. It includes use of sailors (rather than civilian technicians), personnel trained as they would be trained, number of personnel, rank of personnel, etc. The most important element in this striving for realism is

the use of realistic scenarios, including threats, tactics, etc. Our tests are as realistic as possible within the numerous and obvious limitations in trying to run a miniwar. Usually our testing is quasi-realistic -- realistic enough so that the results can be projected to the future, but controlled enough to force encounters and to enable reconstruction for analysis purposes. This tradeoff will be explored later.

Section 2

Analysis Before Project Operations

201. OT&E Planning. With respect to OPTEVFOR, the entire OT&E planning effort is a continuous analysis effort. The beginning is a broad project purpose, then a gross division into areas, a more detailed division in a master plan, still more details in a test plan, finer detail in a series of tests, and finally, within each test, a vernier as to data and measurements. Since steps are interrelated and vary in degree from project to project, the analysis process is best handled as a continuous process, and not divided into master plan, test plan, etc.

a. Purpose. The project purpose is usually stated in broad terms like "... to determine the operational effectiveness and operational suitability of the XXX System." While not stated directly, the following points are implied because they are critical in delineating the purpose: missions, scenarios, threats, fleet criteria, and system functions.

b. Mission. Mission refers to the military objectives related to the tasks to be done. Missions and tasks refer to the battle element that the system being evaluated supports. For a radar antenna, the battle element may be the radar system. For a gun system the battle element may be the ship on which it is installed. For a "whole ship" evaluation, the battle element may be the Task Group itself. Usually a system has different missions and tasks. Some tasks are independent of each other, others may not be. It may be possible to attempt to weigh each mission and combine results analytically. However, this is not recommended.

Each mission is better treated separately in planning and in reporting.

c. Scenarios. A scenario is in effect a sequential description of events occurring during an engagement. It indicates the action during an encounter in the order of its development. The scenario details and directs our interest. It translates the missions and tasks to a workable level. It is fruitful in determining functions, issues, test conditions, etc. The scenarios are key elements in determining project operations.

d. Threat. An important aspect of scenarios is the opposition or threat -- including the type, quantity, capability, and expected reactions opposing our carrying out the assigned tasks. Basically, this is a scenario from the opposition point of view. Be careful to not over- or underestimate the threat and to reference the proper time frame -- when the system will be introduced into the fleet.

e. Fleet Criteria. As we consider the various tasks, we need to establish what we consider to be performance in those tasks. Such criteria should be as general as possible, and should reflect what is necessary in order to be effective, not what is just nice to have. For example, in the FFG 7 ASW task of defending the escorted force, our criterion for judging effectiveness should be: losses to the escorted force. Note that this criterion only states what is important for the task. A second-order criterion in this case would be destruction of the attacking force -- such destruction would be nice, but if we can prevent losses to the protected force in any way possible, we will have successfully

carried out our task. We need to define general criteria for each assigned task. Note that the criteria at this stage do not include numerical goals or specifications.

f. System Functions. Description of the system is formally brought into our procedure. The description is not a technical description -- our interest is not in the different modes, per se, but in the different functions in an operational sense.

g. Preliminary Tactics. We have to know how to use the system under test. Early in OT&E, of course, the tactics can only be preliminary. But the point is that the tactics used during the evaluation directly influence the results. Considerable thought should be given to the tactics to be used -- it may be wise to task some Dev Group to propose the tactics. We should be prepared to defend the tactics used, particularily if the evaluation involves two competing designs.

h. MOMS. MOMS is the probability that the battle element using the system under evaluation will perform its operational mission. As stated paragraph 102.f., MOMS is a combined effectiveness and suitability measure; it is a derived or calculated measure, based on a series of conditional probabilities or functional MOEs, many of which may not be directly measureable during OT&E. Inputs may be necessary from other sources. An illustration of a MOMS is the ASW task for the FFG. The criterion, protection of the escorted force, when translated into measurement terms is: losses in the escorted force; this in turn leads to number of escorted ships lost per attack.

202. Preparation for Planning. This paragraph gives broad

guidance in planning, to indicate the direction of the planning effort. Detailed steps in planning are given in paragraph 215.

a. The general project purpose is divided into warfare areas and then into effectiveness (performance or capability) and suitability (reliability, availability, human factors) objectives. Each of these is best treated separately, at least initially. It is understood that a CEM for performance will be combined with an MOS for availability and a MOMS formed for the final report.

b. A CEM is created for each warfare area based on the various missions and corresponding fleet criteria. Consider the performance for one warfare area, say ASW.

(1) Escort. CEM may be "Probability of a convoy ship surviving a submarine attack."

(2) Patrol. CEM may be "Probability of submarine attack leading to submarine being sunk."

c. For each CEM, the analysis effort proceeds in two structures:

(1) One is to bring the different scenarios and threats into account. This structure aids in test design and in determining test situations during project operations. After project operations, we may find that different scenarios/threats may have different values of the same type of CEM.

(2) The major structure is in determining the type of calculations to be used in deriving the CEM. In a realistic free-play type of operation, we may be able to determine CEM simply; e.g., by a ratio of number of succeses to number of attacks.

More typically we would not have this luxury and CEM must be derived by a formula such as:

$$CEM = P_1 \cdot P_2 \cdot P_3 \cdot P_4 \cdots P_n$$

where P_1 ..... P_n are the conditional probabilities of success in steps leading to the end-objective or subobjective.

d. The type of calculations to be used in determining the CEM, once delineated, leads to subobjectives. For example, if the above series of conditional probabilities are to be used, the subobjectives would be determination of detection capability, acquisition capability, etc.

e. For each subobjective, the dendritic breakdown continues with the determination of the type of data needed to arrive at each probability.

f. Knowing the data need, the next step is to determine the type of measurements and corresponding instrumentation.

203. <u>The Engagement Model</u>

a. The above steps involve a high order of involvement. Early effort in an engagement model can help -- not only in forming subobjectives, but also in other stages of our planning.

b. The engagement model is descriptive of the engagement with the system being evaluated in its battle element based on scenarios, threat, and preliminary tactics.

c. Figure 2-1 gives part of an engagement model for the XXX System under evaluation in its ASW mission. Basically, the modeling is a follow-through of specific events and corresponding outcomes.

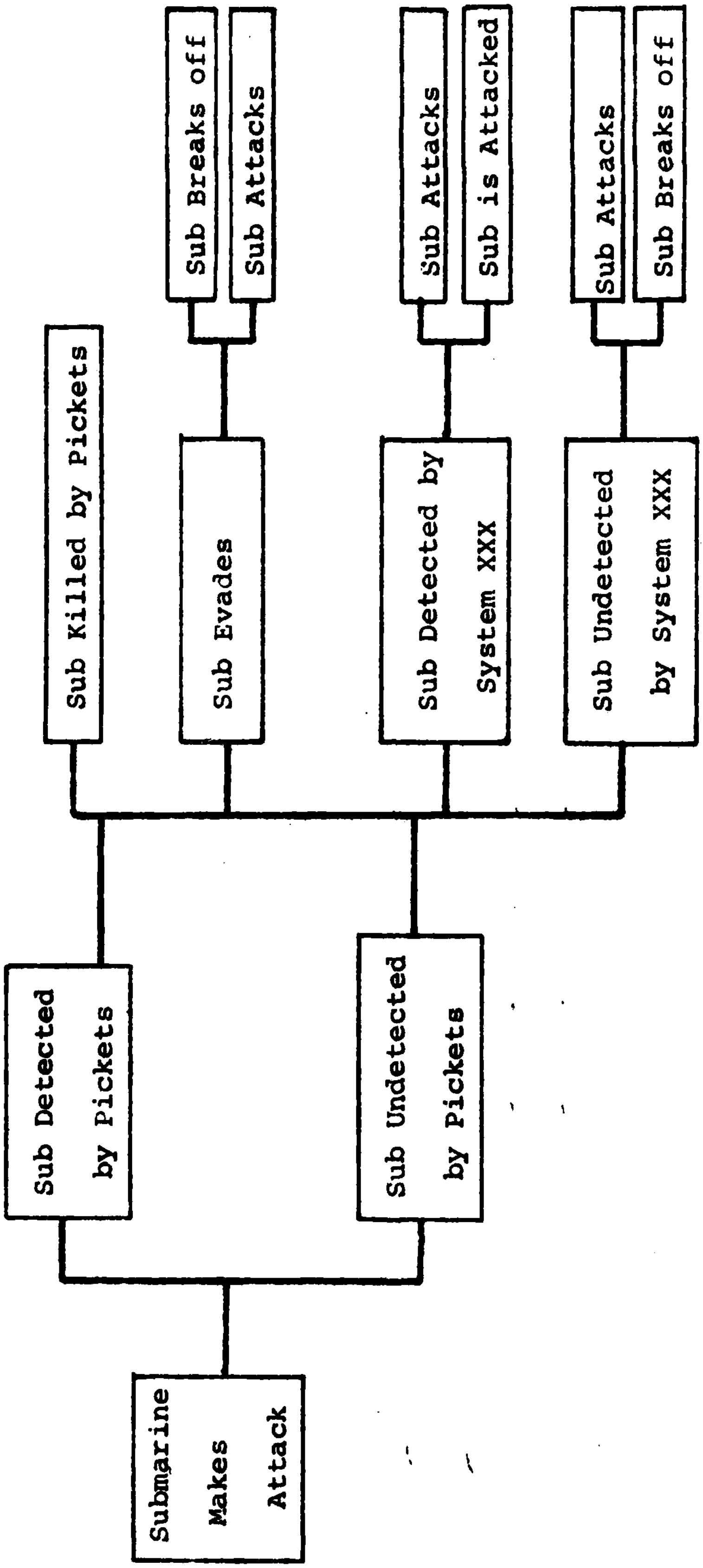


Figure 2-1

<u>Part of an Engagement Model</u>

d. The next step in the engagement modeling is to obtain guesstimates for the likelihood of the various events/outcomes. Some may be based on published reports, some may be group consensus, some may be based on projections of contractual specifications. Some (such as enemy aggressiveness) may be merely guesses that may be handled by a range of values.

e. After the guesstimates are formed, the arithmetic is followed through -- e.g., starting with 100 submarine attacks and ending with some value of the CEM.

f. Sensitivity analysis can also be done, particularly on "sure" guesses and on guesstimates that might be expected to be obtained during project operations. Some critical operational questions can be pinpointed quantitatively at this stage.

204. Uses of the Engagement Model

a. Analysis. The engagement modeling effort can help in the analysis steps.

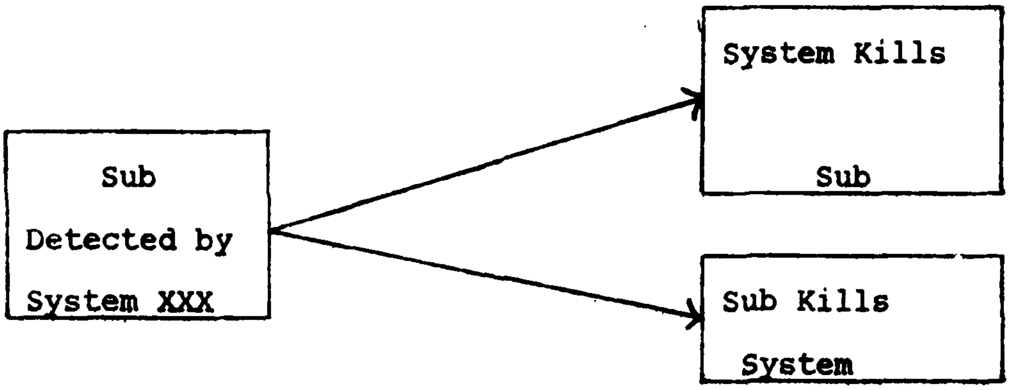
(1) By going through the calculations to determine CEM, we are in a firmer position to specify the type of formula to be used in paragraph 202.c. We can begin to start determination as to the components needing project operations, others already available, etc.

(2) In paragraph 202.d, the modeling effort leads to an obvious subobjective: detection capability for example. Also, the model tells us that some finer subobjectives are in order. For example:

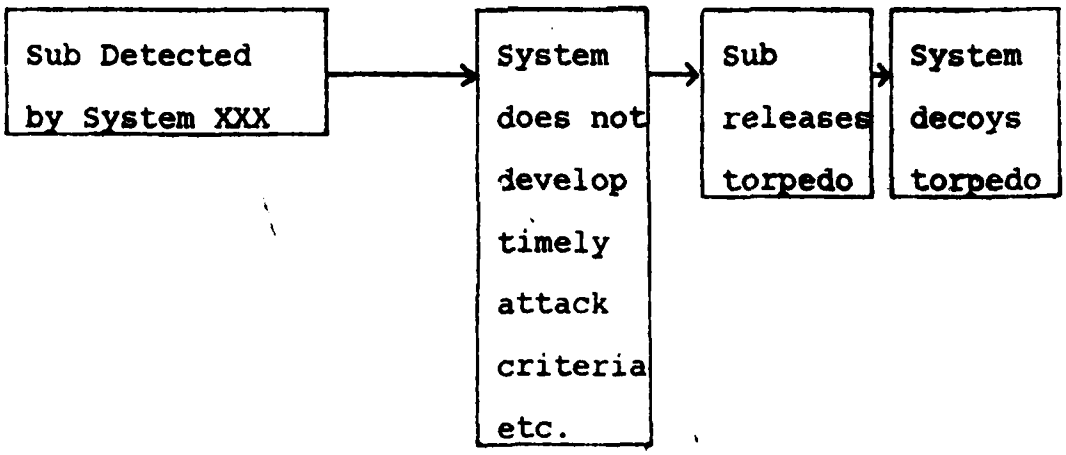
(a) Is the detection capability better when System XXX is alerted by picket forces?

(b) Is the detection capability different when the submarine makes an optimum approach as opposed to after evasion?

A finer cut can be made of parts of the engagement model that in turn can lead to finer subobjectives. For example:



Can be blown up to include



This in turn leads to development of timely attack criteria subobjectives, decoy effectiveness subobjectives, etc. This stepping procedure is continuously recycled and updated as inputs become firmer and firmer. At first, the procedure is done broad-brush. Even in gross terms the procedure is helpful.

b. Long Lead Items. The evaluation procedure may indicate the need for new types of targets, for particular threat intelligence, or for a hybrid-type simulation. The sooner these long lead items are addressed, the more likely they will be available.

c. Allocation of Effort. The CEM calculation will include components that are beyond our test capability -- e.g., probability of warhead damage. The evaluation logic will identify components to be handled by published reports, by technical testing, by simulation of some type, etc.

205. Analysis in the TEMP

a. With respect to analysis, the TEMP (Test and Evaluation Master Plan) gives the logic of the project evaluation viewed as a whole. The TEMP outlines the information needed, the sources to be used for the information including elements to be obtained from at-sea testing, and directly supports its implementation in the Test Plan. Particular attention is given to long-lead items.

b. The parts of the TEMP that give the objectives, missions, scenarios, threats, criteria, employment concepts, etc., are the basis for the analysis effort. The results of the analysis dendritic effort are also given in the TEMP. The culmination at the sub-objective stage may be given as in Figure 2-2. This working figure summarizes the CEM by giving the functional MOEs needed for the CEM. In addition, it depicts the likely sources of the needed information. The information in Figure 2-2 can be presented in other ways. The purpose is to indicate as a concise summary the type of decisions made in the TEMP.

Mission A (ASW)

CEM

	P_{det}	P_{class}	P_{launch}	P_{acq}	P_{home}	P_{hit}	P_{kill}
Scenario 1/Threat 4	F	F	A	B	A	G	I
Scenario 4/Threat 4	F	F	E	E	E	G	I
		MOS					
Scenario 1/Threat 4	H	H	B,D	B,D	B,D	U	N

Mission B (ASMD)

	T_{det}	T_{class}	T_{launch}	T_{fire}	T_{kill}
Scenario 2/Threat 1					

... etc ...

Source Code

A	OPEVAL testing	G	Naval Lab Report
B	TECHEVAL testing	H	3M System
C	TECHEVAL/OPEVAL	I	Limitation to Scope
D	Earlier DT Testing	N	Not Pertinent
E	Simulation	U	Undecided
F	Fleet Reports		

Figure 2-2

Objective Source Matrix

c. Each element in Figure 2-2 is, of course, based on the rationale given elsewhere in the TEMP. For example, P_{kill} given P_{hit} in Figure 2-2 is relegated by the code I to Limitations to Scope. Why the CEM is restricted to Hit rather than to Kill should be explained. For example, simulation is to be used instead of testing at sea for one scenario. The risk of the simulation not being validated should be taken into account. The scope and determination of source, etc., for the elements are the result of tradeoffs taking schedules, cost, and expected resources into account. In addition, while the elements in Figure 2-2 are not finely structured, some fine structure may actually be needed with respect to long-lead items: specially configured targets, simulation, etc.

d. The CEM approach will automatically cover many operational issues. These will pertain mainly to quantifiable issues. However, there are a host of issues that cannot be handled by the CEM approach. Particular effort must be made to include these in the TEMP.

206. <u>Analytical Testing Issues</u>. The analysis effort for the TEMP gives rise to and is strongly influenced by analytical testing issues.

a. Assumptions. The TEMP is based on many assumptions, changes to which may necessitate a review and change in the TEMP. These assumptions include:

(1) Development schedule.

(2) Decision schedule.

(3) At-sea resources; quantity, quality, type.

(4) Target resources; quantity, quality, type.

(5) Instrumentation needs.

(6) Simulation availability.

(7) Engagement model availability.

(8) Documentation availability.

(9) Technical issue resolution.

b. Test Outline. The various operational measures should be defined specifically. The need for definition includes functional events also. (Summary of detection success (MOE) will be determined by the range at the 50% cumulative probability, for example.) The critical parameters relating to system performance should be listed and the expected source of their determination given. The major emphasis would be on the overall test sequence -- that is, which MOE or data would be obtained from factory testing, etc. This outline should note the system integration tests as well as other critical tests such as the software integration tests.

c. Operational Testing. The testing under OPTEVFOR's direct control is amplified in more detail. Definitions, MOEs, scenarios, and threats become more specific. The stress sequence is outlined -- e.g., single target, no CM to multi-target with CM. Differentiation as to "real" targets vice simulators are made. Variational analysis about the key scenarios/threats are included. Preliminary tactics are taken into account. Data requirements including instrumentation accuracy needs and sample sizes are included.

207. <u>Analysis in Test Planning</u>. Analysis in test planning, as an implementation to the TEMP, details how the total resources are to be used to fulfill the project objectives. Many areas involve specific analysis techniques. The Analyst will actually do the methodology; however, many decisions will have to be made in their application. The Analyst is expected to clear the important ones with the OTD. For example, the Analyst may have an allocation of runs at sea so that a particular series will have three replications. What does this sample size mean in terms of confidence? The OTD would be interested in errors such as:

a. Likelihood of reporting operationally important differences that do not actually exist. This type of "false alarm" is called Type I, or Alpha (α) error.

b. Likelihood of not reporting operationally important differences that do actually exist. This type of "false positive" is called Type II, or Beta (β) error.

208. <u>The Function/Variable Chart</u>

a. While analysis techniques will be used by the Analyst, the inputs for these techniques need be generated jointly by the OTD and Analyst. An effective tool for this is the Function/Variable Chart. The Function/Variable Chart is illustrated in Figure 2-3. This example concerns a handheld, heatseeking, surface-to-air missile. The columns list the system functions: the functions in the illustration are detection by radar and/or visual detection, acquisition, gyro uncage, and fire.

b. The rows list the differences among the scenario/threats, including CM and environment. These are called the variables.

Figure 2-3

Illustration of a Function/Variable Chart

Variables	FUNCTION D Radar	V Visual	A Acquisition	G Gyro	F Fire
Mission (Mutual, Self Defense)			X	X	
Ship (PTF, MSO, ATF)	X	X			
Speed (PTF, MSO, ATF)			X		
Evasion (Steady, Evade)		X	X		
Gunners (A/B/C/D, C/E, D/F		X	X	X	X
A/C (F-8, T-28)	X	X	X	X	
Approach (Dive, Strafe)	X	X			
Weather		X			
Sea State			X	X	
			REATTACK		
Mission			X		
Ship/Speed			X (Speed)		
Evasion		X	X		
Gunners		X	X		X
A/C Type		X	X		
Approach					
Weather		X			
Sea State			X	X	
Expected useful runs (%)	90%	85%	60%	55%	50%
Data: Count:	D/N	V/N	A/V	G/A	F/G
Data: Range:	D	D-V	V-A	A-G	F
Data: Error:			A	G	F

The checks indicate which variables are considered important for which function; the checks indicate the test variables by function. The illustration indicates that only three of the nine variables should be tested with respect to radar detection. The Function/Variable Chart also includes other features if available. For example the illustration indicates the expected run size loss. These are guesstimates -- e.g., we may plan for 100 runs, but only 90% may actually be run (and valid). And likewise, because of non-detections, we may only have 60% to determine acquisition. Another useful feature gives the expected amount of data (independent) per run. For example, if two or more weapons were to be used in each attack, the target acquisition phase would double, etc., in sample size. In other words, the number of runs may or may not be the amount of data. This feature is not indicated in the illustration, since only one weapon was to be available.

c. Another useful feature is the type of data to be used in data analysis to determine success or non-success of each function. For example, the illustration gives three types of data for acquisition:

(1) Count. The number of raids having acquisition compared to the number having visual detection.

(2) Range. The range difference from visual detection to acquisition.

(3) Error. The aiming error at acquisition.

d. The Function/Variable Chart gives structure and direction to the discussions between the OTD and the Analyst. As such, it

is an effective summary and documentation. In addition, it gives specific guidance as to data and instrumentation needs during project operations, and to the post-test data analysis -- what data to analyze for what test variables.

e. More important, the chart focuses the discussion on the functional approach that, when services are tight, may lead to other means to satisfy sample size needs. These include combining various tests and scenarios, reattack, dry runs, more than one target or weapon, use of land-based test site data, etc.

f. As a tool for the Analyst, it may aid in combining various scenarios into one matrix or factorial approach. This is an important test design technique for efficient selection of run condition.

209. Supplementing Sample Size. Because fleet services are tight, obtaining sufficient confidence for our report demands a high-level analysis effort -- there are ways to increase confidence other than with a large sample size. These ways include obtaining supplementary data and by test design.

a. Supplementary Data. Supplementary and complementary data or information can be obtained from factory testing, technical testing, land-based testing, simulation testing, etc. The questions of pertinency and validity are critical and are the basis, in part, for the extensive analysis effort described for using the functional approach, etc.

b. Test Design. Test design is analytical jargon for the efficent arrangement or combination of test conditions. Two simple illustrations will be given in the paragraphs to follow.

210. Side-by-Side. If the comparison of interest is between two tactics (or two modes, etc.):

a. We could test at a test range, one week with Tactic A and another week with Tactic B. By extensive instrumentation and data processing and analysis, we can standardize the data for environmental changes, etc., and make the comparison with the standardized data.

b. A more efficient test design would be to test both tactics side-by-side or back-to-back. When we make a run with Tactic A, we immediately repeat the run with Tactic B. Each pair of test runs can be made in different ocean areas, different environmental conditions, etc. Differences in the data within each pair would automatically cancel out the effect of the environment for ocean area. The comparison between tactics could then be made directly with these data differences. See Figure 2-4.

c. Side-by-side testing improves the scope of the comparison (different areas, etc.) with less testing and reduced instrumentation, data processing, and analysis. The efficiency is such that this approach is sometimes used in comparing two pieces of equipment (old versus new) or in source selection problems (manufacturer A versus B). Obviously, we must be able to have both equipments available at the same time -- and the reliability

Figure 2-4

<u>Illustration of Side-by-Side Testing</u>

"Shoot-Out" Sanders vs Magnavox

Run	Date	Sea State	Layer Depth	Company	Detection Range
1	7/3	2	50	Sanders	7500 yds
2	7/3	2	50	Magnavox	8300 yds
3	7/12	6	200	Sanders	2500 yds
4	7/12	6	200	Magnavox	3500 yds
o					o
o					o
o					o
o					o
o					o
21	8/1			Sanders	6600 yds
22	8/1			Maganvox	7200 yds

In Transit, data not available.

Data Difference

(Magnavox minus Sanders)

800 yds

1000 yds

'

'

'

'

'

600 yds

of each should be high. This simple concept can be extended to multiple comparisons of various types.

211. Factorial. If the primary interest is testing various scenarios/threats that cover, for example, various combinations of speed, altitude, CM, and manning conditions:

a. We can, for example, select seven "most critical" combinations and test each of them five times. Each threat result would be based on sample size five. Analytically this test design is simple and straightforward. It is called one-at-a-time.

b. We could, however, run all possible combinations of the various settings as in a factorial. The matrix (illustration) is shown in Figure 2-5. We now have 16 scenarios/threats of which the "seven critical" are included. We can test each of the 16 twice. The total number of runs is similar to the number made one-at-a-time.

(1) A formal analysis technique is available that would permit us to combine the data in various ways. For example, the 32 data runs can be divided into four sets of sample size eight each by speed/altitude settings. See Table 2-1A. Another way would be by CM and manning conditions. See Table 2-1B. Notice that in each table the other conditions combined into each average are balanced -- each average in a set has the same conditions pooled.

(2) Table 2-1C is a work table to simplify the calculations -- it is merely the means of Table 2-1B, but in terms of differences from the grand mean of all the data.

CM	Manning Condition	Speed and Altitude			
		Slow		Fast	
		Low	High	Low	High
Clear	I				
	III				
SOJ	I				
	III				

Note: Each of the 16 test conditions indicated above are tested twice.

Figure 2-5

Illustration of a Factorial Matrix

(3) The result for each threat is found by correspondingly summing Table 2-1A and 2-1C according to the scenario/threat conditions. For example, if scenario/threat A is fast, low, clear, I, we would use 88.2 in Table 2-1A with -3.1 in Table 2-1C -- the answer is 85.1 seconds. All other 15 scenario/threat results can be found in a similar manner.

(4) Notice that even though each scenario/threat was tested twice, these two runs are not used, per se, to obtain each scenario/threat result. Each result is obtained by using means in Table 2-1, each based on a sample size of eight. The confidence in each result compares favorably with those obtained using the one-at-a-time approach, which are based on five runs each. Using about the same number of total tests, the factorial increased the results from seven scenario/threats to 16.

c. There is extensive literature on the factorial to cover various situations. Basically, the larger we can make the factorial matrix, the more efficient -- for large factorials we need not make any repeat runs. In many situations, only half or a quarter of the factorial matrix need be tested (with no repeat runs). See your Analyst for details.

d. The efficiency of the factorial is because each data point is used over and over again in forming the various tables of means. However, each missing data point affects each table. If many are missing, the efficiency is lost. Therefore, the large factorial has a limited use in our work. If many data points are expected to be missing, it is better to use a different test design.

Table 2-1

<u>Illustration of Factorial Analysis</u>

(Reaction Time in Seconds)

A. Table of Means (Eight Data Points Each)

	ALTITUDE	
<u>Speed</u>	<u>Low</u>	<u>High</u>
Slow	66.0	61.1
Fast	88.2	57.5

B. Table of Means (Eight Data Points Each)

	MANNING CONDITION	
CM	I	III
Clear	65.1	61.6
SOJ	58.1	88.0

C. Means: Differences from Overall Mean (68.2)

	MANNING CONDITION	
CM	I	III
Clear	-3.1	-6.6
SOJ	-10.1	+19.8

212. <u>Paper Rehearsal</u>. Experienced OTDs say: <u>The best time to plan the conduct of project operations is after operations are completed</u>. That is, difficulties always arise, wasted effort is hard to avoid, etc. Experienced data analysts say the same with respect to analytical inputs. While we do not have the luxury of hindsight, we try to approach it with a paper rehearsal. After the test scenarios, data requirements, and data analysis techniques are decided, the most critical elements are rehearsed before firming up the Test Plan. This rehearsal can be a simple examination at the blackboard or a complex series of trials on a computer simulation duplicating the planned scenarios. The typical rehearsal includes "creating" data, using Murphy's Law for lost data, etc. See COMOPTEVFORINST 3960.7 for a description of a paper rehearsal and an illustration.

213. <u>Pretesting</u>. The concept of rehearsal can be broadened to include other types of pretesting. For example, if a model or simulation is available, sensitivity studies may give insight into the relative differences among the various scenarios/threats. Based on these studies, project operations may be refined or broadened. An important type of pretest concerns run geometries. For complex geometries, particularily those including countermeasures, a simulated rehearsal may indicate that certain geometries would lead to "no opportunity," and a regroup may be in order. Even if no changes in geometries are made, the run-by-run outputs of the simulation would be a guide to the OTD during the actual project operations.

214. Bias. An important element in confidence is the avoidance of bias creeping into project operations, data analysis, or in the presentation of the report. This is much broader than the OTD maintaining strict objectivity at all times. The operational objective must always be kept in mind.

a. Test Plan. The selection of test scenarios/threats should be based on an expectation of likely use. This expectation can be based on documentation. In all cases, the OTD must use his judgement based on his operational experience. For example, in "shoot-out" type testing to select system A or system B, the scenarios/threats to be tested should be selected so as not to favor either system. While both systems would be tested with identical scenarios/threats, the individual tactics, modes, etc., need not be similar. The tactics to be used with system A should be optimum for it; the tactics for system B should be optimum for it.

b. Test Sequence. The Test Plan should also consider the sequence of testing; extraneous, time-varying effects may distort our results unless special precautions are taken. These effects include environmental changes and the crew learning curve. Special attention should, therefore, be paid to the sequence of testing. Some situations may best be handled by test design. In many situations, some form of randomness in sequence is used. Randomness is sequencing test conditions by a formal procedure using chance. Shuffling cards labeled by condition is the usual method. Randomness is a form of insurance that systematic extraneous trends do not distort our results. As a form of

insurance, there is a premium we have to pay. For example, if we randomize target approach altitudes, the time and fuel needed to vary altitude might materially extend test time or reduce the number of runs. So a tradeoff is involved. Usually, complete randomness is not followed; some stratified randomness is done taking the premium and the risk into account. For example, if the crew's learning curve is in the sharply changing region, we may decide to randomize the sequence of testing conditions. If in this case the premium is too high and test designing cannot help, we may be forced not to randomize and later in reporting comment on the possible distortion.

c. Qualitative Data. Qualitative data are usually suspect with respect to bias because of the many ways bias can creep in with questionaires, ranking methods, etc. However, at times, important points can only be backed-up with qualitiative data. In addition, discussion of results and operational implications may be based in large part on opinion. The point is that automatic discounting need not be done. We should report the basis for credence in such opinions. These include qualification of the source or sources. We would not want the opinion of the OTD discounted if he observed the project operations and others present shared his opinion, and he successfully defended his opinion in the review of the report.

d. Reporting. There is a thin line between good reporting and biased reporting. Some results may be given stronger emphasis than others. Some findings may be considered minor and not even

mentioned. The numerous decisions made in presentation should be based on operational considerations. Even so, the needs of various groups using the report may differ. One facet of reporting -- credibility -- deserves emphasis. Not only should our report be accurate, but it should also include credibility. For this reason, non-standard analysis techniques are described and an annex usually includes a run-by-run data summary.

215. Steps in Test Planning. Test planning is too complex to be done "by the numbers." However, steps will be given, to indicate the approach. Early in the planning stage, perhaps for TEMP preparation, the methodology given in paragraphs 216-219 is begun. This methodology has inputs: information on missions, scenarios, threats, and tactics. Early in the project, this information may be tentative -- becoming more firm as time goes on. These inputs are the primary responsibility of CNO and the OTD. The analyst as a team member may comment, interpret, etc., but these inputs are not analytical per se. The responsibilities of the analyst in the steps given below are related to his areas of expertise. See Section 6. The methodology is two-fold:

a. The MOMS, CEM, and MOS approaches outlined below are used for quantitative elements. Paragraphs 216 and 217 lists the steps for this phase.

b. The other approach which is used in parallel is the "critical issue" approach using the EEA technique for qualitative elements. See paragraph 218. Paragraph 219 repeats many steps given earlier, but from a "conduct of test" point of view to indicate concurrency of these various paragraphs.

216. <u>Methodology Steps for Quantitative Elements</u>

STEP

1. Treat each mission/scenario separately.
2. Consider the test system as part of the larger system (determine inputs/outputs) it is supporting.
3. Verbalize the CEM.
4. Use previous reports to refine the CEM. (These reports are also useful in many of the following steps.)
5. Refine the conditional part of the CEM.
6. Include a survival clause in the CEM.
7. Precisely define the CEM.
8. Define each event in the CEM.
9. Determine the functional MOEs in the CEM
10. Determine the procedure or formula for using the functional MOEs to form the CEM.
11. Form the general test objectives using the CEMs.
12. Form the subobjectives by using the functional MOEs.
13. Plan conduct of test around scenarios (see paragraph 219).
14. Form the function/variable chart.
15. Form the test design.
16. Determine data needed for the functional MOEs.
17. Determine data to be observed during test.
18. Determine other data sources, particularily simulation.
19. Check amount of services, number of runs, and test sequence.
20. Outline data analysis.

21. Include a statistical confidence procedure.

22. Include qualitative modifications from paragraph 218.

23. Give important limitations.

24. Run rehearsal and sensitivity studies on simulation or with paper/pencil.

25. After project operations, the methodology steps are done in reverse order for analysis.

26. Individual functional MOEs per mission are derived. (Based on Step 16)

27. The particular functional MOEs are combined for all scenarios if statistically valid.

28. The CEMs are found (Using Step 10) per mission.

29. The various values of CEM are listed by environment, threats, etc. per mission.

30. The CEM values are modified by the qualitative elements in Steps 22 and 23 and the EEA approach.

31. The quantitative values of the CEM and qualitative modifications are reported as operational effectiveness.

32. Repeat above for MOS (Paragraph 217).

33. Combined estimate of effectiveness and suitability is made.

217. Determing Suitability Objectives

STEP

1. Select a mission and scenarios.
2. Verbalize MOS.
3. Precisely define MOS.
4. Define each event in MOS.
5. Define MOS in terms of functional parameters.
6. Relate functional parameters to specific hardware.
7. Determine mission time per function per hardware.
8. Determine method or formula for combining functional parameters.
9. Form general test objectives using MOSs.
10. Form subobjectives using functional parameters.
11. Review current version of project operations plan.
12. Determine functional parameter data to be available from project operations.
13. Determine functional parameter data to be supplied from other sources
14. Determine if project operations need revision.
15. Determine sample size needs.
16. Relate to expected services.
17. Outline data analysis procedures.
18. Repeat, if necessary, with other missions.
19. Relate above to reliability model, if available.
20. Plan to use such a model extensively if one is available.

218. Methodology Steps for Qualitative Elements

STEPS

1. Select a mission.
2. Review source documents for critical questions.
3. Group and list critical questions.
4. Supplement with operational questions.
5. Each question is handled separately.
6. Repeat above for other important missions.
7. The resultant matrix due to (5) and (6) is listed.
8. Redundancies are eliminated.
9. Collate list with topics from functional MOEs.
10. Note specific subobjectives in common (paragraph 217).
11. Eliminate specifics better handled quantitatively.
12. Note specifics better handled qualitatively.
13. Note specifics only to be handled qualitatively.
14. Each resultant specific is considered a test objective.
15. List sub-elements for each test and form subobjectives.
16. Further subdivision may be necessary in few cases.
17. If a subobjective is specific enough, it is termed an EEA.
18. Checklist questions are formed to answer each EEA. These are descriptive, answered by yes/no, a word, or number.

19. After project operations the checklist questions are answered individually, and then combined into answering subobjectives.

20. Test objectives are answered combining quantitative with qualitative measures.

219. Determining Project Operations

STEP

1. Select a mission
2. Select key scenarios.
3. Expand description, action events in scenarios.
4. Determine key elements in scenarios.
5. Use scenarios as basis for project operation.
6. Determine elements to be simulated.
7. Verbalize (initial cut) scenario/project operations.
8. Correlate with CEM and functional MOEs.
9. Form MOE function/variable chart.
10. Determine MOE data needed for functional MOEs.
11. Determine MOE data to be available from project operations.
12. Determine MOE data not to be available from project operations.
13. Determine MOE data to be obtained from other sources.
14. Form test design.
15. Determine sample size needs.
16. Relate to expected services.
17. Outline data analysis procedures.
18. Run rehearsal.
19. Modify approach if necessary.
20. Repeat above with another mission.
21. Eliminate redundancy.

Section 3

Analysis During Project Operations

301. Introduction. One of the objectives of a Test Plan is to document the handling of the numerous decisions that must be made during project operations. During test planning, these decisions can be considered in the quiet of the office, and the OTD is thus freed to conduct project operations unhampered by numerous questions. However, many problems cannot be foreseen, some decisions can only be made on-scene because prior information was lacking, and some decisions have to be modified because of the on-scene situation. Regardless of the amount of planning and pretesting, the conduct of project operations is seldom straightforward. The way projects operations are actually conducted obviously affects the data analysis and the final report. Analysis-type thinking, then, has an important part in many decisions made on-scene. In some evaluations, the OTD may have the Analysts close by or on-scene. In other projects, the OTD must apply his own analytical background and experience.

302. On-Scene Preparations

a. On-scene, the OTD quickly realizes that his Test Plan needs updating. Many changes can occur. The system under evaluation may not be complete, or a component may be down because of an on-order part. Run geometries may have to be modified because of a substitute target, etc. The technical certification is incomplete. The above indicate major decisions, and obviously, the analytical thinking used in planning will help.

b. Other decisions will be mundane and would have little impact. Some, seemingly mundane, may have important repercussions. For example, should COMEX also include erasing non-test targets being routinely tracked by the test equipment? To be operational, we should not erase; however, later reconstruction/data processing efforts may be increased many-fold if we do not erase.

c. An important preparation tool is pretesting or dry runs. Dry runs are rehearsal runs, made in transit or in situ, usually as complete as possible except for missile fire or use of other limited expendable items. Dry runs are used to debug the control and logistics of the operations. Particular emphasis is put on the data instrumentation process. This is a final check that all non-test areas (such as range control, data takers, etc.) do not waste runs.

d. Formal briefings to the test units involved and informal discussions should stress the objective of the testing, primarily that it is an evaluation. The aim is not to make the system look good or poor; system results do not reflect on the unit operating it. If this is stressed, individual biases are minimized.

303. Analysis During Project Operations

a. The OTD's aim, of course, is to follow and complete the Test Plan. However, surprises may occur and decisions have to be made. It is important to recognize when a decision has to be made; the tools used in planning can help. For example, simulation results in sensitivity testing may have indicated the importance of environment on detection. Or the rehearsal runs made on a land-based test site may overlay the runs at sea. A

particularly difficult decision must be faced following a series of failures. Suppose a series of sonar trials begins with a number of no-detection runs. Should we continue or regroup? The answer depends on many elements, most of them obvious. For example, statistical variation in data is expected. Even if repeat runs are made, wide fluctuations are not surprising. Statistical formulae can help. On the other hand, a technical examination may indicate the need for corrective action, and the problem is clearly not statistical but technical. Therefore, a technical review of the operational runs should be made continuously. If possible this should be done on-scene.

b. The OTD should attempt to keep the structure of project operations constant during the trials. Care should be taken in varying extraneous (non-tested) elements. For example, in ASMD the Evaluator may want to change his procedure in the middle of the test. If the change is made, with or without the concurrence of the OTD, the runs affected should be noted. This note may be an important guide in later analysis of reaction times. This illustrates the importance of the OTD's Notebook.

c. The OTD should review each run continuously or daily. Data sheets or complete quick-look printouts should be collected and examined. Missing data or unclear remarks should be tracked down. Differences in actual from planned scenarios should be noted and explained if possible.

d. A running account of progress should be made. This includes not only the runs made and still to be made, but also the results of each run. This running chart or check sheet is useful

for sequential-type decisions as to sufficiency, priority, and perhaps, reallocation. Specific analysis techniques can be used if they are prepared for in advance. For example, a sequential decision chart can be prepared in the planning stage for later use at sea. Each run result at sea is plotted; when the results fall in a "stop test" region of the chart, the OTD has sufficient confidence to reallocate his remaining resources to other objectives. The more typical situation is in the closing days of project operations with insufficient resources left to complete the planned runs. If priorities are not helpful, this is a difficult decision area. Generally speaking, we would want to test as many of different scenarios or planned test conditions as possible before using resources to make repeat runs.

Section 4

Analysis After Project Operations

401. Introduction. The basic objective of analysis after project operations is to quantify the MOS and CEM for scenarios/threats correspondingly found homogeneous, and to indicate the degree of confidence in the reported quantities. As stated earlier, the analysis effort before project operations forges the trail or roadway; post-operations analysis uses the same roadway in the opposite direction. So the general approach has already been covered, but will be repeated here. Before analysis or synthesis begins, some important steps must be taken:

a. Review the pre-operations analysis effort and Test Plan to determine the impact of project operations. For example, a particular scenario may have only been tested once instead of a planned number of times. This scenario may be relegated to a demonstration and excluded from data analysis.

b. Review current fleet needs to determine possible changes in emphasis in analysis. A change of emphasis may be dictated by results during project operations. For example, if the system was obviously ineffective during project operations, we may want to analyze some areas (causes) in greater depth than planned.

c. Review reports of technical testing for insights into possible relationships or for possible use of technical results and/or data.

402. Steps

a. The measurements taken during project operations and the subsequent data processing are the basis for analysis. Experience

indicates that this area is filled with unexpected problems. Some types of data may not be recoverable. Decisions have to be made at various states. These decisions include a dropping of some types of data because of a lack of timeliness or because of questionable validity of certain runs. Since data processing is the beginning of confidence or lack of confidence (which cannot always be quantified) in our final results, it is an inherent part of the analysis. The end of data processing is a run-by-run data summary that includes each run made or attempted during project operations. This run summary gives the actual test scenario/threats or test conditions, environmental conditions, the various performance data, validity codes, and remarks. Not only does this run summary provide the formal base of subsequent analysis, it will usually be an annex in the final report.

b. The next series of steps are basically data analysis or statistical analysis. First, the run-by-run data summary is scanned to update the analysis techniques outlined in the Test Plan. Some columns of data may not be amenable to formal statistical analysis; some may be too sparsely filled, others may be practically constant. For example, if one column contains only three data points, it would be better to report it as such and not mislead the reader by statistical elegance. Or if sea state varied only minutely, the analysis time would be better spent on other variables.

c. The first formal step in statistical analysis is to measure and adjust (standardize) for extraneous effects such as changes in environment, effects of practice, etc. Formal statistical

techniques are used to determine if a relationship does exist, what the relationship is, and the importance of the relationship. This is the basis for standardizing (or normalizing) the data for further analysis. This information is of value, per se, and should be included in the final report. Note that when a technical relationship is available (inverse square law, etc), it should not be used unless the statistical results agree; the assumptions involved with the law may not be valid in the project operations.

d. The next step is to determine the degree of synthesis; how should the data for different test scenarios be combined, if at all. The approach is to start with the smallest element, a single test scenario or test condition, and combine it with as many others as is valid. The limitations to the combining of different sets of data are determined by formal statistical techniques. If these techniques indicate significant differences, then the data would not be pooled. For example, suppose in project operations the classification was correct in 75 of 100 runs against a diesel target and in 25 of 100 runs against a nuclear target. Pooling would give 100 correct classifications out of 200 runs (50%). This 50% would be misleading; too high in one case (nuclear), and too low in another (diesel). If statistical techniques indicate significant differences among results for different scenarios, separate values for MOEs, CEMs, and MOSs are indicated in subsequent analysis steps. Thus, in the above illustration, separate CEM values would be determined for the two types of targets. (In addition, an overall CEM can

be determined by weighing the two separate CEMs by intelligence estimates of the distribution of the two target types.)

e. The above synthesis pertains to combining various test situations, but with each type of data analyzed separately -- detection range data are analyzed separately from classification data and from hit/miss data, etc. The next synthesis pertains to combining various types of results into functional MOEs. For example, detection range results are combined with "no contact" results and the MOE, probability of detection, by range result is formed. Or the reaction time for radar detection is combined with threat analysis time and with decoy launch time, etc., into a total reaction time MOE.

f. The next step is to combine MOEs into CEM, using MOEs determined from project operations and MOEs from other sources, if appropriate.

g. The CEM is then combined with the MOS results of availability analysis and the MOMS value is determined. The actual combining is perhaps by formulae or computer engagement model. If the MOMS values are reported, the derivation used is reported, as well as some indication of confidence.

h. The next step is to determine the level of confidence in our final results. Statistical confidence can be obtained for some components used in the CEM and MOOS. However, there may be other components with unknown confidence. Some may be based on extrapolations, some may be based on intelligence estimates, etc. And, of course, there are important aspects that are qualitative.

Sensitivity analysis may help in indicating the impact of possible error in some functional MOE components such as intelligence estimates. Our confidence in some component values may be so small that the scope of the CEM is restricted and only a partial CEM is reported. Our confidence in the CEM/MOS values may be so small that the decision may be made not to report these values as absolute but only relative, <u>if at all</u>. Confidence in the CEM values, as do all confidence problems, involves a series of tradeoffs. On one hand we realize that the formation of MOMS is based on modeling, and the more we model, the more the danger of large errors. On the other hand, if we eliminate this type of error and avoid the MOMS, we end up with a series of functional MOEs. This forces each reader/decision-maker to integrate (mentally) the details into a personal judgment. <u>We are not, then, doing the complete project evaluation</u>.

Section 5

Suitability Considerations

501. Scenario Approach. The analysis thinking already presented in general terms applies to suitability also. Scenarios, operational realism, test planning, etc., are as important in suitability as in effectiveness.

a. For example, different scenarios may result in different suitability measures. If a system uses different hardware in different modes, the suitability result for one scenario may be different from another. Suppose different failure rates pertain to search mode hardware, track/acquisition hardware, and launch/firing hardware. The suitability measures would vary strongly by scenario if mode times vary by scenario. A complex system with different modes/hardware has many suitability measures.

b. The scenario approach to suitability requires an imaginative operational interpretation of the criteria. For example, an ESM processor may have an MTBF criterion of 470 hours. This technical or design criterion, while it is interesting, has little operational meaning. A more meaningful operational specification is derived by considering the planned fleet use of the processor. The processor is to be used with a LAMPS helo: 90-day deployment, 40 sorties per month, a typical sortie being 1.5 hours. No repairs are to be made at sea except direct replacement. Only one spare is available. Evaluating operational suitability requires taking the above into account in forming the MOS. An operational measure is having the processor (or spare) functioning without materiel abort during the entire deployment. The OTD recommends

setting the threshold probabilities at 0.95 -- in other words, 95% of wartime deployments should be completed with the processor functioning at the end of the deployment. This operational criterion obviously goes beyond the simple technical measure of MTBF.

c. With respect to suitability, certain simplifications in analysis are possible. In many projects, different test conditions need not be imposed for suitability determinations. For example, a radar is tested, materiel-wise, whether an actual target is present or not. The stress on the radar is practically the same. In many projects, then, data on suitability are accumulated while test conditions are being varied for effectiveness purposes.

d. Seldom can the scenario approach be tested directly or completely. However, standard analysis techniques can be used with hardware failure rates and repair rates determined from at-sea testing to derive the results expected in various scenarios. The at-sea testing is necessary to determine the hardware reliability and then a reliability model is made to take the different scenario-dependent hardware and modes into account. Thus, an observed hardware failure at sea may be critical in one scenario but minor in another.

502. Operational Suitability

a. Operational suitability encompasses a spectrum of elements, broadly classified as reliability, maintainability, supportability, etc. Each of these broad categories contains subelements, some of which can be quantified, and some of which cannot. The significance

of reliability in hardware design, which can be quantified in various ways, and the intangible characteristics of operator proficiency, which currently defies quantification, represent the opposite ends of this overall spectrum.

b. In order to determine the quantitative measures of operational suitability, data are accumulated to determine MTBF (mean time between failure), MTTR (mean time to repair), MTFL (mean time to fault locate), MTBM (mean time between maintenance), A_O (operational availability), MSI (maintenance support index), R (mission reliability), MMT (mean maintenance time), etc. The numerical values of these indicies are not the only indication of operational suitability, but they are needed to relate the at-sea observations to the operational scenarios.

503. Data Collection And Processing

a. Failure analysis, an important step in data collection, is determining not only what failed but also why. Failures are catagorized as follows:

(1) Assignable

(a) Design

(b) Manufacturing (quality control).

(c) Personnel (maintenance, handling).

(d) Non-hardware (software).

(e) Outside envelope (stress, test conditions).

(f) Combination of above.

(g) Exercise (non-valid).

(2) Unassignable (unexplained, random)

The process of deciding the catagory of failure is difficult and

may lead to disagreement with the DA or contractor. OPTEVFOR has a fleet-assessment point of view. For example, a failure may or may not be deemed to have occurred in a realistic fleet situation (handling damage is in this category). Some catagories may be ignored for some reliability indices, but not for others. In other words each failure must be examined critically and analysis must be carefully done.

b. Quantitative data relative to DOWNTIME should be collected on standard 3M, MAF, or NAMP forms. Some modification to the instructions for completing these forms will be required to ensure that all data are recorded. Quantitative data relative to UPTIME should normally be collected in an Operator's Log, Debriefing Log, etc. Extreme care must be taken in establishing the requirement for this log to ensure that it does not impact on the performance of the operator. Automatic data recording, if available, takes precedence, and portable voice and video tape recorders should be considered.

c. Qualitative data consist of comments entered on data collection forms plus questionnaires and debriefs following project operations. In some cases, these qualitative data may be partially quantified to the extent of identifying the number, background, etc., of respondents.

504. Reliability. "The probability that an item will perform its intended function for a specified interval under stated conditions." Operationally it is usually expressed as a probability of completing an engagement without a failure.

a. Because systems under test are frequently complex and contain many components or electronic circuits, operating time under full stress is the important measurement. Excluding wear-in and wear-out, the failure rate is usually stable. The steady state situation, constant failure rate, leads to the use of the exponential distribution. This is used as:

$$R = e^{-\frac{t}{MTBF}}$$

where R = probability of completing an engagement of time t without a failure.

b. MTBF in the above equation is the operational test time divided by the number of failures. How the failures are counted influences the operational interpretation of R. For example, if MTBF is based on counting only failures that cause mission-abort, then R is the probability of completing an engagement without mission-aborting failure. The observed MTBF is usually considered as a best estimate from which confidence estimates or predictions can be made using statistical techniques. MTBF is usually determined by dividing operating time by the number of critical (mission-aborting), and major (mission-degrading) failures. Minor failures are usually ignored in determining operational reliability (and for determining most maintenance measures).

c. In many projects, reliability pertains to a combination of software and hardware. From an analysis point-of-view it is best to separate software and hardware failures, and later combine both into a combined, total picture. The primary reason for initial

eparation of software and hardware is related to maintenance nd logistics. It may take seconds to "restore" a software failure, hile it may take hours to correct a hardware failure. While eparation is desirable, it may be difficult to pin down a failure s software. To avoid misinterpretation we use the expression hardware failures" and "non-hardware failures."

d. Failure is not always dependent on operating time. Reliability of one-shot devices (where the outcome of a test of the evice can be classified as a success or failure) is measured as he proportion of success to total number of trials.

e. Reliability evaluations may also contain a qualitative ssessment regarding any aspects of design, workmanship, installaion, or operations that affect reliability.

f. In projects where the system under test has more than ne mode of operation or mission, it is important to report the eliability of each mode or mission rather than one overall figure. ata must be carefully analyzed to insure correct application to parate modes or missions. Complex systems may require other rms of analysis such as modeling or failure rate weighing to count for a utilization factor that differs from the one perienced in testing.

5. Maintainability

a. An assessment of the effort or work required to keep a stem in a state of readiness during a deployment, the maintainility assessment can be expressed in many ways, with most of e common measures using time:

(1) MTTR (mean time to repair).

(2) MTFL (mean time to fault locate).

(3) MTBPM (mean time between preventive maintenance).

(4) MMHPFH (mean man-hours per flight hour).

(5) MSI (maintenance support index).

Most of these classic measures should be translated to operational terms. For example, using the length of a deployment, demands on the system, MTBF, and MTTR, one can make a statement such as "During a typical 30-day deployment, the system will be undergoing maintenance 16% of the time."

b. Failures trigger the time measurements discussed above. Operationally, failures trigger another action--the effort to overcome the impact of the failure and continue the engagement. The time it takes to shift to a backup system, our secondary mode of operation, in order to complete the mission is not necessarily related to the time it takes for actual repairs, and it should be noted. For example, a blown circuit in a computer power supply may take a few seconds to restore, but it may take hours to regenerate the automatic display of tactical information.

c. Maintainability evaluations also contain a qualitative assessment regarding aspects of design, installation, or operation that affect maintenance. In some cases, systems perform without a failure or with few failures during an evaluation -- yielding little or no data on maintainability. To ensure that data are obtained, always make provisions for a backup maintainability demonstration (using prefaulted modules for example) so that the maintainability assessment need not be entirely qualitative.

d. The purpose of calculating MTTR is to measure the corrective maintenance time involved in restoring to a fully operational status a system that has failed or is significantly degraded. Corrective maintenance time is the time during which one or more personnel are working on a critical or major failure to effect a repair. Corrective maintenance time includes: preparation, fault location, part procurement from local (on-board) sources, fault correction, adjustment/calibration, and follow-up checkout times. It excludes supply and administrative time. In calculating MTFL, the purpose is to arrive at a measure of the difficulty of troubleshooting the equipment.

e. Experience has indicated that repair times and fault locate times are not statistically normal. A few repair actions take an extremely long time compared to the bulk of the actions. An arithmetic mean of such data is not representative. Statistically this is handled by calculating the geometric mean. The geometric mean is more useful in summarizing, predicting, and estimating such times as repair and fault locate. The geometric mean is obtained by averaging the logarithms of the time data and then finding the antilog.

f. Another significant index of maintainability is MSI, a measure of the number of man-hours of active maintenance time (preventive and corrective) required for each hour of equipment operating time.

g. For systems with built-in test equipment, the maintainability assessment is not complete without an evaluation of the adequacy of this maintenance tool. Some useful numerical indices

are false alarm rate, degree of coverage, and percentage of correct isolations. If data are not available, a qualitative assessment of the test equipment may be made through questionnaires and interviews with maintenance personnel concerning the usefulness, adequacy, and the confidence they have in the output of the equipment.

506. Availability

a. A_o (operational availability) best expresses the probability that the weapon system will be ready when needed to engage the enemy. During a deployment, the interval of time during which the system may be called upon to carry out an engagement is called engagement demand time. A system is available during this period if it is operating, in standby, or off and can be brought on-line within an acceptable delay. A system is not available if it is "hard down," that is being repaired or in preventive maintenance and cannot be restored to operation in time to meet the threat.

b. A_o is expressed as the percentage of engagement demand time during which the system can engage. This quantity is the availability that is most meaningful to a Fleet Commander and is what may be expected in an operational environment. A_o takes into account all maintenance actions as well as delays awaiting procurement of repair parts. Modifications to A_o may exclude standby or off-time from the calculations.

507. Logistics Supportability. To satisfy the OT&E requirement to assess logistics, it is not sufficient to determine that logistics support will be in accordance with some established

Navy-wide system. As an OTD, you are in a position to forestall the totally unsatisfactory situation of introducing a weapon system to the fleet that is not supportable at the time of its introduction.

a. Systems designated for test and evaluation and installed in a fleet unit are usually supported by a package of spares assembled by the manufacturer. This package does not necessarily represent the on-board spares of the ultimate installation. Use of the spares from the package assembled by the manufacturer must be carefully monitored. Further, the type and number of spares actually used must be compared to the APL (Allowance Parts List) if available.

b. Several qualitative areas should be examined when assessing logistics supportability. These included:

(1) The availability and adequacy of special test equipment, tools, and/or handling equipment.

(2) The need for continued contractor support.

(3) The requirement for any special test or maintenance facilities.

c. Software requires support, too. As with mechanical systems, problems (program errors) can be reported by message; unlike mechanical systems, repair (a program change) can be transmitted by message too. If the weapons system involves software, include software support of in your assessment.

d. LFA (logistics factor of availability) is a numerical measure of the quality of logistics support. LFA is defined as:

$$LFA = 1 - \frac{D_s}{U + D_s + D_o}$$

where D_s is the time spent awaiting delivery of spares from beyond the unit

D_o is all other down time

U is up time or demand usage time

LFA is the fraction of time that we are not waiting for external spares. Conversely, 1 - LFA is the fraction of time that was occupied by logistic delay. LFA is so defined that if

$$A_o = \frac{U}{U+D_s+D_o}$$

and A_1 is the availability with perfect logistics, i.e:

$$A_1 = \frac{U}{U+D_o}$$

then

$$A_o = (A_1) \quad (LFA)$$

e. The availability, adequacy, and accuracy of all technical information required for use and maintenance of the system under test is evaluated under the heading of technical documentation. Required technical documentation usually includes operating and maintenance manuals and PMS documentation including MIPs and MRCs. Current procurement policies do not require final documentation prior to the production decision. However, some form of maintenance manual will be needed to support an OPEVAL.

The manuals should be assessed to the extent possible with respect to accuracy and usefulness to the operator and maintenance technician. Inaccuracies or lack of required information should be reported. Preliminary technician manuals should be evaluated using the standards required of the final manuals. If draft or preliminary MRCs are available, they should be evaluated for applicability and accuracy as if they were the final issue. In some new systems, the number of operating modes and combinations of modes may represent a marked departure from those previously available. The question of how to set the system up for optimum performance under different environmental conditions is an important one. When dealing with such a system, the Developing Agency should be tasked with providing an Operating Guideline, which is usually based on technical information and theory. Inaccuracies in the technical information or Operating Guideline are evaluated under Technical Documentation; recommendations for better methods to employ the system fall logically in the Tactics category.

508. Compatibility. Compatibility includes the effects of the environment on the equipment and effects of the equipment on the environment. When addressing compatibility, the system can be evaluated from these points of view: physical, functional, and electrical/electronic/acoustic. The following outlines what to look for.

a. Physical environment considerations include the effect of such factors as:

(1) Climate and weather conditions including wind, temperature and humidity.

(2) Motion in various sea states or heavy air turbulance.

(3) Maneuvers at high speeds and at critical speeds.

(4) High "G" loading or altitude extremes for aircraft systems.

(5) Depth and pressure extremes for underwater systems.

(6) Sea water temperature, pressure, and solidity.

(7) Icing and saltwater.

(8) Shock and vibration from machinery, gunfire, or ship speed.

(9) Ecology requirements.

b. Functional environment considerations include the effect of such factors as:

(1) Size and weight including ancillary units and cabling.

(2) Handling and stowage requirements, including those for accessories.

(3) Ambient room temperature, ventilation, exhaust, and air supply requirements.

(4) Internal equipment air or water cooling requirements.

(5) Mechanical and electrical interfaces and interconnections with other equipments.

(6) Energy requirements such as power, voltage, or fuel.

c. Electrical/electronic/acoustic environment considerations include the extent and effect of such factors as:

(1) Electrical, radio frequency, and acoustic interference to and from other systems in the test ship, or weapon.

(2) Main power supply variations including the effect of power failure and line frequency and voltage instabilities including voltage extremes and transients.

509. Interoperability

a. Factors to consider include:

(1) The ability to transfer information with negligible distortion.

(2) Proper impedance matching, bandwidth, and data rates.

(3) Fluid flow and mechanical linkages.

(4) Electrical and mechanical loading.

(5) Any degradation of performance of the ship or aircraft or its other systems that result from the installation or use of the system being evaluated.

b. Interoperability also concerns operation between the system being evaluated on "own ship" and other ships, aircraft, or shore stations with which it must operate. The interoperability point of view is large and includes such additional considerations as command and control and mutual interfaces (such as when radars or sonars on different ships are operating on the same frequency). For some projects, interoperability will include the capability of operating with elements of the other services; the Air Force, Army, Coast Guard and the Marine Corps.

c. Interoperability also concerns human engineering and human factors, i.e., man/machine interface. See your Analyst for this important area.

Section 6

Support

601. Guidelines. This chapter delineates the various areas of support, both internal and external to OPTEVFOR. The OTD should delegate as much effort as possible in other expertise areas to free him to work in his own areas of expertise: operational aspects and team leadership. This includes delegation to the Analyst in the areas that the Analyst has expertise.

602. Missions/Threats/Scenarios/Criteria. As the program develops, details will become firmer. Parts will sometimes be missing. As the lack becomes critical to the evaluation, OPTEVFOR may have to request such guidelines formally. If the gaps delay the "critical path," the OTD has no recourse except to use his best judgment in filling up the gaps.

603. Instrumentation/Data Processing. OPTEVFOR usually makes a serious attempt not to get involved with the actual instrumentation suite or data processing, per se. Those areas are usually delegated to a Naval Laboratory or contractor. OPTEVFOR should take particular pains to monitor such efforts; however, OPTEVFOR's responsibility is basically to determine:

a. What data we need, including event and axis definitions.

b. How much data per run we need.

c. How accurately we need the data.

d. In what form we need the data.

e. When we need the end-product: the run-by-run data spread sheets.

Our experience in these areas (see paragraph 302) indicates the value of early dry runs to "proof" instrumentation in time for corrective actions.

604. Simulation/Gaming/Modeling

a. There is a spectrum of simulations, varying from paper studies to at sea testing. Figure 6-1 gives this spectrum and corresponding uses by OPTEVFOR. This is a long-lead area. OPTEVFOR's responsibility is early identification of the extent of need and to recommend ways to meet the need.

b. Besides the Program Manager, there are many groups within the Navy having specific knowledge and expertise that ought to be tapped by OPTEVFOR. In particular, OP-96, CNA, and ONR should be checkpoints for every system program in which OPTEVFOR becomes involved. They frequently have conducted studies directly applicable to the system in question, or can direct us to the agencies who have. (Conversely, for their studies they often have need of system and operational parameters that could be furnished by us.)

c. Even if a simulation is available, contacts should be made early. Modifications may have to be made to make it more operational, etc. In the modeling area, OPTEVFOR takes particular pains to monitor such modeling rather than become technical matter experts.

d. In the actual use of the simulation, specific tests are given in the Test Plan. These test may include:

(1) Sensitivity studies.

USES Early Involvement	Experience, Studies, Paper/ Pencil	Simulation: Effectiveness Gaming, Engagement Models	Simulation: System w/wo actual hard-ware w/wo man in loop	Simulation: Displays, man in loop mock-ups, trainer	Full System: Ashore	Full System: At-Sea/Air
System Feasibility	X	X				
Critical Issues	X	X				
Operational Assessment		X	X			
Reliability Assessment			X*		X	
Human Factors Assessment				X	X	
MOE Study			X		X	
OPEVAL						
Parameter Sensitivity	X		X		X	
Test Plan Rehearsal			X			
Diagnosis of test Abnormalities			X			
Augmentation of Test Runs			X		X	
Extrapolation of Conditions	X		X		X	
Effectiveness Measures			X		X	X
Users Handbook			X		X	X
Tactics Development	X	X			X	X
Reliability Assessment			X*			X
Human Factors Assessment	X	X	X		X	X

*Reliability modeling.

Figure 6-1

(2) Rehearsal of project operations.

(3) Validation of simulation.

(4) Augmentation: project operations.

(5) Real threats, real areas.

(6) Countermeasures.

(7) Measure of mission success.

(8) Tactics Improvement.

605. Project Analysis

a. The OTD is expected to use his Analyst before, during, and after project operations. As a vital team member, the Analyst has many functions to perform. Some "cost" is involved in becoming familiar with the project, i.e,. the Analyst has to be thoroughly briefed before he "earns" his keep. Table 6-1 includes the element of familiarization and gives the analytical support functions that the OTD can expect from his Analyst.

b. The Analyst is completely responsible for very few functions per se. Note all the "...aid the ..." in Table 6-1. This is to denote his being a team member. In each function, however, the Analyst is professionally responsible for all analytical and statistical techniques. In addition, he, as a team member, is professionally responsible for all analytical steps, including timeliness.

c. The Analyst is in danger of becoming an assistant OTD. For example, certain sections of plans and reports are best prepared initially by an Analyst. However, the OTD should insure clarity, understanding, etc. To meet deadlines, the Analyst would be tempted to prepare more and more of the plan and report.

The Analyst should avoid preparation of briefings, becoming a format or report expert, etc.

d. The OPTEVFOR Analyst is expected to service many projects. This automatically means outside support on major evaluations; delegating as much of the routine to others as possible, while using extreme care in spelling out the specific support needs. The plan for data analysis must be in more detail than usual, and most critical is continuous monitoring of the effort at all stages.

e. The OPTEVFOR Analyst is responsible for project work done by other analysts for OPTEVFOR. This includes early determination that outside support is needed, the amount and type needed, the group best fitted, and close monitoring at each and every stage.

606. Types of Support Augmentation. On large complex system evaluations the question is how better to obtain the necessary support: technical agency or contractors. Frankly the selection is not too critical; we have seen excellent work done by both groups. Other things being equal, the following guidelines apply:

a. Technical Agency. When the support work can best be done at the source of the inputs, away from the Headquarters/Squadrons/Detachment.

(1) We may have to work closely with the Technical Agency in the planning stage.

(2) The project operations are continuous rather than intermittent, affording little opportunity for changing test objectives.

(3) There is extensive data processing.

(4) Technical analysis and support is necessary.

(5) The data analysis procedures can be adequately reseen.

(6) Minor slippages in deadlines can be tolerated.

b. Contractors. When the support work can best be done the Headquarters/Squadron/Detachment.

(1) Numerous technical agencies may be involved.

(2) The situation is fluid, leading to changes in jectives.

(3) Data processing is minimal or done by a third rty.

(4) Technical analysis and support is minimal or too verse.

(5) The data analysis procedures cannot be adequately reseen.

(6) Direct control is necessary to insure concentrated, ll-time effort to meet deadlines:

Table 6-1

Generalized Functions of an Analyst in a General Project

Familiarization

1. Become familiar with project, system, Navy's requirements.
2. Become familiar with conceptual design studies.
3. Become familiar with simulations, test beds.
4. Become familiar with current status, calendar of events.
5. Become familiar with software management plan.

Project Logic

6. Aid OTD in preparing scenarios.
7. Aid OTD in determining or interpreting criteria.
8. Aid OTD in determining CEM and MOS.
9. Aid OTD in the dendritic structure of objectives.
10. Aid OTD in developing TEMP.

Long-Lead Time Aspects

11. Aid OTD in determining fleet services (preliminary).
12. Aid OTD in identifying new target needs, obtaining threat information, having simulations available.
13. Aid OTD in determining instrumentation needs, analytical support (preliminary).
14. Aid OTD in special funding requests.

Early Involvement

15. Become familiar with manufacturer's plan, data, results.
16. Aid OTD in identifying operational issues.
17. Aid OTD in resolving operational issues.
18. Aid OTD in making early assessments.

OPEVAL Planning

19. Aid OTD in firming-up OPEVAL test objectives.
20. Aid OTD in firming-up OPEVAL test allocations, test conditions, sample sizes.
21. Aid OTD in firming-up OPEVAL data needs.
22. Aid OTD in firming-up analytical support requirements.
23. Conduct sensitivity studies.
24. Review DA test plan.
25. Aid OTD in merging DA/COMOPTEVFOR needs, allocations.
26. Review expected services.
27. Aid OTD in drafting Test Plan.
28. Conduct rehearsal of plan.
29. Attend cut board on plan, help revise plan.
30. Advise on impact of changes during at-sea operations.

After At-Sea Operations

31. Conduct, analyze, report simulation validation.
32. Review TECHEVAL data processing.
33. Monitor or conduct OPEVAL data processing.
34. Monitor or conduct OPEVAL data analysis.
35. Monitor or conduct simulation tests.
36. Review TECHEVAL report.
37. Monitor or conduct effectiveness and reliability modeling.
38. Prepare analysis results for OTD.
39. Aid OTD in DSARC briefings/message reports.
40. Aid OTD in preparing operating doctrine, tactics.
41. Aid OTD in defending report.

Section 7

Summary

701. Purpose of OT&E. The primary purpose of OT&E is to estimate and predict the prospective system's operational effectiveness and operational suitability. This estimation is a projection of the test results to some future time when the system will be used in the fleet against real targets. A meaningful projection requires testing in as realistic a manner as possible in:

a. Missions and scenarios using expected tactics.

b. Targets that strike back.

c. Production-type hardware.

d. Typical personnel in rate, training, and number.

e. Intended operating environment.

702. Realism. To be realistic, we should include as broad a scope in our evaluation as possible. It is more efficient, more valid, and more useful to test as many different scenarios or conditions as possible before repeating any. There are many constraints to complete realism. For example, in early OT&E, hardware may be breadbroad, tactics information may be incomplete, etc. Test services including number of firings, targets, etc. are so limited that free-play must be controlled so encounters are forced. Another typical constraint is instrumentation. We cannot afford complete realism because it prevents knowing what went on, if we had a valid opportunity, etc. If we cannot reconstruct, then we are playing and not evaluating.

703. Test Data In theory, the operational MOE in evaluating the kill potential of a weapon system is the hit/miss count. This is

ideal if sufficient firings are available. However, test services are always limited. In our work, hit/miss counts and other direct test data must be amplified by:

a. Technical, engineering examination of each firing.

b. Extensive use of supplementary data from non-firing tests, from functional component tests, etc.

c. Use of modeling and computer simulations.

d. Use of expert judgement

704. Analysis in OT&E

a. In OT&E, the stress is on evaluation as well as on test. The lack of complete realism coupled with limited test services limit the impact of direct testing and focus more need for evaluation. It is this evaluation process, that, hampered by numerous constraints, necessitates analysis. This is the basis for intense interest and the need for a high order of analysis. The objectives of analysis are to insure that:

(1) In planning, we ask the correct questions.

(2) In testing, we use the minimum amount of services.

(3) In analysis, we answer the questions correctly.

b. Analysis in OT&E draws heavily on operations research and statistics. Analysis is more an art than a science using scientific procedures and techniques. With the stress on operations, only selected scientific procedures are useful. Even these must be adapted. In all cases, the requirements must be expressed or defined in operational terms. Even DOD standard terminology must be interpreted in operational terms.

c. An evaluation requires a thorough, sometimes intensive, analysis effort. The OTD has managerial responsibility in this area -- he is not responsible for the specifics of the analysis. The Analyst is responsible to provide the analytical approach, specifics of test design and planning, professional techniques used, and data analysis. The Analyst may have to direct others in doing some of this, such as in data analysis, but he is responsible for the output. The Analyst can only function with inputs and guidelines supplied by the OTD. The best modus operandi is to work as a team, starting with an initial discussion of the mission, scenarios, etc. Then each does what he can do better because of his training, experience, etc.

d. Data analysis first determines test conditions that are similar in terms of results. These form homogeneous groups. Data within a group may vary only to a minor extent, while from group to group the data may be quite different. A cardinal sin in data analysis is to use an overall average of different groups, based on the number of data points obtained during project operations. The proper way to obtain an overall average is to weigh each group average with the relative occurrence expected in realistic use of the system. If such weights cannot be obtained, then it's best to report each group result separately.

e. The skipper on the bridge is only interested in whether his firing hit or missed the target. If it missed, he doesn't care that it missed due to software or hardware failure or poor performance. So why not just count the hits and misses, and

that's our analysis? If our sample sizes were large, this would be sufficient. However to squeeze as much information as possible from our small sample sizes requires detailed analysis. Initially, it usually is more fruitful to determine effects separately. (Stresses are different, mechanisms of failure vary, etc.) Then, the separate results are combined based on the scenarios to answer evaluation criteria or what the skipper on the bridge wants to know.

705. MOE Approach. The MOE approach is the basic analysis approach. In OT&E, this is extended to include the support principle. The critical MOE includes more than the system being evaluated; it includes that which the system is supporting. For example, in evaluating a new sonar, the sonar is viewed in terms of the system it is supporting. For a certain mission, the sonar may support the fire control system. The critical MOE would be the timeliness and accuracy of torpedo launch parameters. Suppose testing indicated poor fire control solutions. Even if the sonar per se was technically fine and the interface was what lead to the poor solution, the sonar would be deemed not operationally effective in this mission because a submarine CO would rather have the old system than the new. Technically the interface may not be subject to test; it may not be the responsibility of the sonar program manager. However, we are not evaluating NAVMAT or the program manager. This extension of the MOE is an important difference between developmental and operational viewpoints.

706. Operational Effectiveness and Operational Suitability

a. The likelihood of the system being "up" is a major factor in operational suitability. This factor is a combination of availability and reliability.

(1) Availability pertains to the weapon system being ready when needed. (Is the aircraft on deck "up" materiel-wise? Is it ready to start its mission?)

(2) Reliability pertains to the weapon system completing its mission without a materiel abort. (Does the aircraft complete its bombing mission without a materiel abort?) Stress and repair help differentiate between the two. Reliability is important with full stress when no repairs are possible during the stress -- e.g., a missile in flight. Availability is important when repairs are possible -- e.g. a continuously operating radar. While the likelihood of completing a deployment period without a failure (reliability) may be of interest, this measure is not so important since the radar can be repaired.

b. Operational effectiveness pertains to system performance when the system is "up," i.e., no mission-aborting failures.

c. While effectiveness and suitability are determined separately, the evaluation is not complete until both are combined into an overall measure. This, MOMS, includes:

(1) The probability of a system being up when called on.

(2) The probability of the system remaining up throughout the mission.

(3) The effectiveness of the up system in performing its mission.

Thus, operational questions like ". . .was the target killed . . ." are answered by calculation rather than by an observed count of hits and misses. With small sample sizes the calculation method is preferred.

d. Analytically speaking, we should not penalize a "bad" run twice. That is, if a missile misses the target because of effectiveness or a hardware failure, it should be counted as a miss in effectiveness <u>or</u> reliability but not both. Again speaking analytically, effectiveness and suitability are similar. The principles, practices, techniques, etc. that apply to one apply to the other.

707. <u>System-Level Testing</u>. Operational evaluation includes as much of the entire system as possible. For example, missile evaluations should include some warhead firings against "real" targets. Tradeoff analyses usually indicate that very few warhead firings are worthwhile considering the loss of targets, loss of exercise data, etc. Regardless, experience has indicated the value of at least one complete system check.

708. <u>Mission Orientation in Testing</u>. The stress on operational measures in paragraph 705 requires more than event-by-event or one-on-one testing. Testing should relate to at least a whole engagement, and ofttimes to a deployment or a blue/gold cruise. For example, suppose a new helo was being evaluated in an amphibious operation. Rather than having the MOMS pertain to having the helo complete one sortie (ship to shore to ship), the measure

should include completing the required number of sorties to conduct the amphibious mission with a typical number of helos.

709. Responsibilities in OT&E

a. COMOPTEVFOR is the Navy's only OT&E agency; thus we are the only ones to plan and report on OT&E. While others may conduct the project operations according to our test plan, we are in charge. OT&E is our responsibility.

b. While we are the Navy's experts in how to test operationally, certain inputs that are critical to operational testing and evaluation should be determined for us by CNO. These include missions, scenarios, preliminary tactics, expected threats, fleet needs, and operational requirements and criteria.

c. Working with these inputs, operational analysis can be most beneficial in OT-0 and OT-I to ensure military usefulness in the final product. This has two facets:

(1) Insurance as to military, operational, and combat inputs early in system development.

(2) Risk estimations and determination as to the success of the development at each phase beginning with OT-I. Growth expectations, risk methodology, etc. are included. The analysis effort for the TEMP is critical. The analysis logic of test phasing, use of long-lead items such as digital simulation, MOMS statements, etc., all have far-reaching effects.

d. COMOPTEVFOR is charged with being independent. It is obvious why we should be free of DA ties. This independence also pertains to the user, the fleet. We must refrain from being "in bed" with the fleet. We must be objective in our appraisal.

The user would always want a proposed system that is an improvement. However, trade-off analysis may indicate that the improvement is marginal compared to other proposed systems, etc.

710. The Stress in OT&E

a. The stress in OT&E is correctly focused on combat, on the hot war situation. While this is the primary consideration, the following should also be an important part of the operational criteria:

(1) Cold War. If cold war functions are different from those of a hot war (usually less), this difference should be taken into account. The criteria (e.g., for detection, acquisition, and classification) should reflect a double use: cold as well as hot war.

(2) Fleet Firings. The evaluation of an advanced torpedo stressed the fact that there would be weekly training firings when the system was introduced into the fleet. Retrieval, refurbishment, etc. became critical evaluation aspects.

b. We should not be content with preparing for the next conflict by reliving the last. Warfare is basically a history of change. If we put all our eggs in one basket, we are playing a dangerous game. The concept of a weapon mix is a useful consideration in our approach.

c. Reduced manning is the modus operandi in the Navy -- only in extreme circumstances will a prospective system be deemed effective and suitable if it requires an increase in ship manning. The reality of reduced manning must be stressed in our evaluations.

d. Certain elements are not pertinent to OT&E. For example, design or contract specifications are used by the DA as the basis for contract compliance and for certification of readiness for OPEVAL -- they are not criteria in OT&E. Cost analysis in a formal sense is not part of OT&E -- equipments are cost-examined during concept formulation and contract definition, not during OT&E.

e. OT&E has as an end product a report. This report must be accurate, valid, timely, etc. -- and creditable. Not only must we be right, but we must also convince others that we are right.

(1) Whatever we do must be carefully documented.

(2) Measurements must be quantitative if at all possible. However, quantitative measurements are seldom sufficient -- they must usually be supplemented by qualitative elements based on operational experience. Qualitative results are the weakest link in the evaluation chain. To strengthen this link we use a high degree of structure: specific questionnaires in a test plan, debriefings, etc. More important, we broaden the base of our results as much as possible. "In the opinion of the OTD" has less impact that "all qualified on-scence observers, including representatives of the DA and the contractor agree..."

f. OPTEVFOR's involvement with acquisition programs has indicated that most programs tend toward success-orientation, with overly optimistic schedules and procurement plans. Software development is often a major problem area; experience has shown that at least a third of the effort in software development will be in testing and correcting system integration deficiencies. If this effort is not budgeted, the evaluation will be in difficulty.

To determine the degree of success-orientation in a program, the program structure and the following should be examined:

(1) Software Management Plan.

(2) Human Engineering Program Plan.

(3) Integrated Logistics Support Plan.

(4) Training Plan.

Reviews of these plans in early stages of a program can be of great benefit.

Section 8

Glossary Of Special Analytical Terms

Accuracy. Refers to the deviation of a result from the "true" value. In a broad sense, accuracy refers to the validity of our report in predicting fleet results. This is not the same as precision.

ANOVA (Analysis of Variance). A technique useful in analyzing multi-variable, multi-setting factorials.

Bayesian. Refers to a formal incorporation of prior information, to reduce the amount of testing. The validity of this approach is still questioned by many analysts.

Binomial. The jargon for hit/miss, yes/no count-type data with two categories.

By-Directional. A method of testing in sets of two runs, the second of the pair geometrically opposite in direction from the first of the pair, to balance out effects of wind, current, etc.

CEM (Combat Effectiveness Measure). A measure of effectiveness specifically pertaining to system performance in completing its mission assuming the hardware is "up".

CEP (Circular Error Probable). A summary measure of bombing or tracking accuracy defined as a circle, centered at the aimpoint or mean point of impact, to include 50% of the data. The median of radial errors is one of many ways to determine CEP.

Chi Square (χ^2) Distribution. A particular type of sampling distribution. Tabular values are useful in analyzing count-type data.

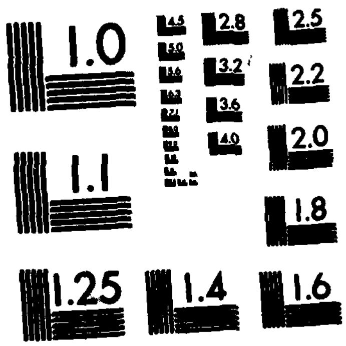
AD-A124 194 PROJECT ANALYSIS GUIDE(U) OPERATIONAL TEST AND EVALUATION FORCE NORFOLK VA 05 FEB 81 COMOPTEVFOR-INST-3960.8 2/2

UNCLASSIFIED F/G 5/2 NL

END

FILMED

DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

Conditional Probability. The probability that an event will be successful, given that the previous event in the functional process or chain has occurred.

Confidence Coefficient. The chance that a confidence interval has of including the true value.

Confidence Interval. An interval that has a designated chance (the confidence coefficient) of including the true value.

Confidence Limits. The end points of a confidence interval.

Confounding. Refers to results that cannot be attributed to a particular single variable or cause.

Contingency Tables. An analytical technique useful to determine significant effects with count-type data.

Correlation Coefficient. An analytical measure of how well changes in one variable are concurred in with changes in another variable. Usually denoted by the symbol r.

Correlation Index. The square of the correlation coefficient (r^2). The index gives the ratio of the variation explained by the independent variable to the total variation in the dependent variable. This is a measure of efficiency of fit.

Data. Refers to the basic outputs of data processing that are subject to analysis. Data may be continuous, quantitative type or count (hit/miss) type.

Degrees of Freedom. In data analysis, usually sample size less one or the number of test settings less one.

DEP (Deflection Error Probable). A measure of bombing or tracking accuracy in deflection, it is the interval centered at the aimpoint or mean point that will include 50% of the errors in the

deflection axis.

Dependent Variable. The test data or variable affected by the test or independent variables.

Error of the First Kind. The alpha error (α) of rejecting a good system. Attributing a difference to results when in reality there is no difference. Also called "Type I" error.

Error of the Second Kind. The beta error (β) of accepting a poor system. The chance of missing an important difference in results. Also called "Type II" error.

F Distribution. A particular type of sampling distribution. Tabular values are useful in the analysis of variance technique.

Factorial Experiment. A test design (matrix) in which settings of each test variable are tested with all settings of every other variable and combinations.

Fractional. A test design in which only - selected partial of a full factorial matrix is tested, depending on the absence of certain interactions.

Function/Variable Chart. An analytical process to determine which variables should be tested by which function in a system evaluation.

Geometric Mean. The mean of a set of data that has been transformed to logarithms, the mean of the logs found, and the anti-log found of this result. Useful in analysis of detection ranges, reaction time, or repair time data.

Histogram. A bar diagram representing a frequency distribution.

Independent Variables. Causal variables, usually controlled at certain test settings. These are the variables that affect the dependent variables, the test data.

Interaction. The tendency for the test data to be dependent on the combination of two or more variables and to give a result different from the sum of the individual contributions.

Lateral Range. A summary measure of performance obtained by first transcribing the area of frequency of occurrences into a step (all or none) function. The value at the step is the lateral range. Sweep width is twice this range.

Latin Square. A test design useful when three test variables do not interact.

Least Squares Method. A method of fitting a line that minimizes the sum of squares of deviations from the fitted result. Used in regression.

Mean (μ, $\bar{X}$). A popular measure of central tendency, the centroid of a frequency distribution. A parameter in the normal distribution. Also called the arithmetic mean or average.

Mean Square. Measures the quantitative importance of each effect in analysis of variance. Also the variance or square of the standard deviation.

Measurements. Refers to the large quantity of positional, etc., measurements taken that are combined with other sets by data processing into test data.

Median. A measure of central tendency defined as the value of the middle datum when the data are monotonically arranged.

MOE (Measure of Effectiveness). A numerical measure of how well a task is done or an objective is met, a generic term.

MOMS (Measure of Mission Success). A measure that combines measures of performance and suitability.

MOS (Measure of Operational Suitability). A measure usually pertaining to system availability, reliability, and maintainability in completing its mission.

Non-Parametric. An analytical approach that does not assume a particular distribution as a basis for the analysis techniques.

Normal. A type of probability distribution that is usually due to a multitude of variables, each small in effect. Can be logarithmic normal when effects are relative. Also called Gaussian.

Normal Probability Paper. A type of graph paper scaled so that a cumulative normal distribution will plot as a straight line.

Null Hypothesis. A tentative hypothesis that there is no difference among conditions, which is then tested analytically for significance.

One-at-a-Time Approach. A test design that first studies one variable completely, then another, etc., in sequence. Interest centers on a few particular conditions that cannot be integrated.

One-Tail Test. A test of significance when the alternate to the null hypothesis includes only one critical value, either less than or more than, but not both as in the two-tail test.

Operating Characteristic Curve. The curve that gives the probability of acceptance as a function of sample size and the true value of true answer.

Performance Measure. A summary measure based on combining all the individual data available for the same type, i.e., dependent variable.

Population. An analysis concept representing the totality or universe of test conditions, pieces of equipment, fleet operators, etc., that are sampled during project operations.

Precision. A measure of the variation of the data among themselves, usually as determined by the standard deviation among the data from the mean value. The smaller the standard deviation, the better the precision.

Probability Distribution. A distribution of relative frequencies (based on large samples).

Radial Error. Refers to miss distance when the data are in terms of absolute distance from the aimpoint without regard to bearing.

Randomness. An intuitive concept referring to an approach that leads to disorder and unpredictability of individual data. Useful in determining the sequence of testing to minimize such effects as environment changes, practices, equipment, wearout.

Regression. An analysis technique to fit a line or curve or hyperplane to a set of data. The technique includes determining the coefficients and correlation testing for significance, etc.

REP (Range Error Probable). A measure of bombing or tracking accuracy in the range direction. It is the interval centered at the aimpoint or mean point that will include 50% of the errors in the range axis.

Repeatability. A measure of precision to include variations because of short-term effects and measurement errors.

Replicatibility. A measure of precision to include as much variation in time, area, etc., as our testing permits.

Replication. A complete test of all test conditions before retesting any conditions.

Reproducibility. A measure of precision to include variations from ship to ship, target to target, etc., usually important but unattainable in our evaluations.

Sample. The number of objects, operators, and conditions that were actually observed to represent the totality of the population of objects, operators, and conditions.

Sequential Approach. A test design used in conjunction with a chart that delineates possible stoppage of testing after each run.

Sets Approach. Refers to testing all comparisons of interest once in a group before proceeding to a repeat of the groups. Also called side-by-aide or back-to-back when only two comparisons are involved.

Side-by-Side. Refers to testing two items during the same time frame to minimize external effects on the comparison. Also called pairing or back-to-back.

Significance Level. A selected chance value (α) or statistical threshold used with test of significance to decide whether or not two means are similar or significantly different.

Standard Deviation (0.5). A parameter in a normal distribution measuring precision.

Stepwise Regression. A computerized curve-fitting program that selects only the significant variables from the input model.

t-Distribution. A probability distribution useful in tests of significance between two means.

Test Condition. The controlled test trial with each test variable held at a particular setting.

Test Design. An analytical combination of test variables and test settings to form a group of test conditions or experimental design. Also includes the sequence of testing and sample size determinations.

Tests of Significance. An analysis technique to determine whether the observed difference in means can reasonably (as set by the significance level) be attributed to chance variation in the data.

Two-Tail Test. A test of significance when the alternate to the null hypothesis includes two critical values (less than or greater than). This is contrasted with the one-tail test.

Truncated Data. Refers to data of a continuous type that has a minimum or maximum because of turnaway, etc. Also called censored data.

Variance. The value of the square of the standard deviation or mean square.

END